When the Money Runs Out:

Do Cash Transfers Have Sustained Effects on Human Capital Accumulation?

March 30, 2019

Sarah Baird
Department of Global Health
Milken Institute School of Public Health
George Washington University
950 New Hampshire Ave. NW
Washington, DC 20052
sbaird@gwu.edu
202-994-0270

Craig McIntosh
School of Global Policy and Strategy
University of California, San Diego
9500 Gilman Dr.
La Jolla, CA 92093-0519
ctmcintosh@ucsd.edu
858-822-1125

Berk Özler
Development Economics Research Group
The World Bank
1818 H Street NW, Mail Stop MC3-306
Washington, DC 20433.
Phone: 202-458-5861
Email: bozler@worldbank.org

When the Money Runs Out:

Do Cash Transfers Have Sustained Effects on Human Capital Accumulation?*

Sarah Baird, Craig McIntosh, and Berk Özler

March 30, 2019

Abstract

The five-year evaluation of a cash transfer program targeted to adolescent females points to both the promise and limitations of cash transfers for persistent welfare gains. Conditional cash transfers produced sustained improvements in education and fertility for initially out-of-school females but caused no detectable gains in other outcomes. Significant declines in HIV prevalence, pregnancy and early marriage observed during the program among recipients of unconditional cash transfers (UCTs) evaporated quickly after the cessation of support. However, children born to UCT beneficiaries during the program had significantly higher height-for-age z-scores at follow-up pointing to the potential importance of cash during critical periods.

Keywords: Cash Transfers, Medium-term Impacts, Human capital, Adolescents

JEL Codes: C93, I15, I21, I38, J12, J13

* Sarah Baird (corresponding author), email: sbaird@gwu.edu, George Washington University; Craig McIntosh, email: ctmcintosh@ucsd.edu, University of California, San Diego; Berk Özler, email: bozler@worldbank.org, The World Bank. We thank seminar and conference participants at the Center for the Study of African Economies, Columbia University, CU Denver, Georgetown University, IFPRI, Mannheim, McGill University, Middlebury College, Monash Development Workshop, Oregon State University, Otago Development Workshop, Otago International Health Research Network, Labor Econometrics Workshop, PopPov Annual Research Conference, University of Maryland, University of Oklahoma, University of Oregon, University of Southern California, University of California, Berkeley, University of California, Los Angeles, Washington Area Development Economics Symposium, Yale University, and the World Bank. We thank everyone who provided this project with great fieldwork and research assistance and are too numerous to list individually. We gratefully acknowledge funding from the Global Development Network, the Bill and Melinda Gates Foundation, 3ie, NBER Africa Project, and the World Bank. Ethical review committees at the National Health Sciences Research Council (Malawi, Protocol #569) the University of California at San Diego (USA, Protocol #090378), and George Washington University (USA, Protocol #061037) approved the study design. The trial is registered at AEA RCT Registry (#AEARCTR-0000036). The findings, interpretations, and conclusions expressed are entirely those of the authors.

1. INTRODUCTION

The past decade has witnessed an impressive growth in the number, volume, and types of cash transfer programs in developing countries. A rigorous evidence base has shown that cash transfers can have significant effects on household consumption and educational attainment in the short-run, even if the poor receive these transfers with few strings attached (Fiszbein et al. 2009; Baird et al. 2013; Haushofer and Shapiro 2016; Garcia and Saavedra 2017; and Molina-Millán et al. 2018a). Furthermore, several working papers and journal articles that have been published in the past couple of years point to some robust evidence of longer-term effects on schooling, but mixed findings on other important outcomes, such as skills, employment, and earnings (Molina-Millán et al. 2018a, for a review). It is still an open question whether such programs can improve the wellbeing of their young beneficiaries when they become adults – after the cessation of support. This question is particularly pertinent for Conditional Cash Transfer (CCT) programs, which are built on the premise that they not only fight current poverty, but also promote human capital accumulation for the next generation. As cash transfer programs continue to grow as major vehicles for social protection, it is important to understand if these programs break the cycle of intergenerational poverty, or whether the benefits simply evaporate when the money runs out.[1]

This paper fits into this growing literature and attempts to contribute to it in two important ways. First, it presents experimental estimates of the impact of a two-year CCT program targeted at adolescent females in Malawi more than two years after the program ended (or four to five years after baseline) – using both a pure control group that never received treatment and another treatment arm that was offered equal-sized *unconditional* cash transfers (UCT). Second, it uses data on a rich set of outcomes (education, childbearing and marriage, health, labor market outcomes, empowerment, and subjective wellbeing) for the target population of young females, as well as information about children born to them during the study period.[2] The resulting analysis is

[1] There is a recent wave of transfer programs, generally conducted by NGOs, which aim to lift households out of poverty using larger lump-sum transfers during a limited period of support (Bandiera et al. 2017a; Banerjee et al. 2015; Haushofer and Shapiro 2016, Haushofer and Shapiro 2018). Evaluations of these programs are generally concerned with current poverty reduction rather than human capital accumulation among children. As such, while the question of sustained effects is also pertinent for these studies, they are less relevant for our examination of longer-term impacts on adolescent beneficiaries.

[2] We also collected data on the husbands of married respondents, but chose to exclude that analysis in this paper based on previous feedback from reviewers. Given that there is significant attrition in our husband data and many of the adolescents are still transitioning into marriage (only 40% of the baseline schoolgirls are married two years after the program compared with 81% of baseline dropouts), the detailed secondary analysis required goes beyond the scope of this paper. For preliminary analysis of these data, the reader can refer to Baird, McIntosh and Özler (2016).

a comprehensive assessment of the relative effects of CCTs and UCTs targeted to adolescents for two years during a period of transition into adulthood – conducted more than two years after they stopped receiving the transfers.[3]

For any intervention to have a sustained effect, it needs to lead to an increase in the stock of some asset that produces a stream of returns in the future, i.e. some accumulation of capital – whether it takes the form of human, physical, or social capital. However, the causal pathway from program implementation to final outcomes can be circuitous. For example, a program that provides cash grants to groups of unemployed youth for income generating activities may have lasting effects on earnings through the accumulation of physical (productive assets) and human (vocational skills) capital (Blattman, Fiala and Martinez 2014). Alternatively, large unconditional cash grants to poor households may increase future earnings by increasing investments in productive assets, such as livestock (Haushofer and Shapiro 2016).[4] Still, small monthly cash transfers over a long period of time may lead to increased consumption after beneficiaries exit the program by increasing savings and investments in small-scale agriculture or non-farm activity (Gertler, Martinez and Rubio-Codina 2012, Handa et al. 2018) or by stimulating entrepreneurial activity (Bianchi and Bobba 2013).

For programs targeting younger people, the causal pathway to improved welfare over the long run is more likely to be human capital accumulation, either in the form of education and skills or health – especially reproductive and sexual health for adolescent females. Even when young women attain higher schooling and delay fertility and marriage, low quality education, credit constraints, and low demand for skilled labor can stunt income gains. Without economic

[3] Cash transfers during adolescence may be particularly effective as this is a critical period to expand one's capabilities by investing in human capital. In fact, adolescent girls are viewed as a key demographic target group to successfully break the intergenerational transmission of poverty in developing countries (Levine et al. 2008). Unfortunately, for many boys and girls in developing countries, adolescence entails a fleeting transition from childhood to adulthood, when they are suddenly expected to "behave as adults even though they are not biologically, cognitively, or emotionally ready to assume adult responsibilities" (Naudeau, Hasan and Bakilana 2015). Adolescent females in particular face a multitude of hazards – ranging from school dropout, to child marriage and teen pregnancy, to physical and mental health problems, to gender based violence (Baird and Özler 2016). Young people's capabilities and functioning (Heckman and Corbin 2016) during this period not only have immediate consequences to their own lives, but also longer-term benefits to their offspring and communities at large (Lloyd and Young 2009; Duflo 2012). Interventions that help adolescent girls reach their full potential by increasing their education, improving their skills, and delaying childbearing have the potential to create a virtuous cycle that improves health, especially child health, and women's empowerment – ultimately leading to higher economic growth (Canning, Raja and Yazbeck 2015).

[4] It should be noted that, despite promising impacts in the shorter-run, longer-term evaluations of both of these programs indicated convergence between the treatment and control groups with respect to employment, earnings, and consumption (Haushofer and Shapiro 2018; Blattman, Fiala, and Martinez 2018).

independence, potential gains in women's agency, intra-household bargaining power, and empowerment are foregone.

Programs targeted to adolescent females may not only delay marriage and childbearing but may also benefit the development of their own children. A distinct and mostly U.S.-based literature, largely using quasi-experimental methods, has examined the very long-term effects of being exposed to cash, 'near cash,' or other safety net programs during childhood (e.g. Currie and Almond 2011; Aizer et al. 2016; Hoynes, Schanzenbach and Almond 2016; Chetty, Hendren and Katz 2016) and has demonstrated beneficial effects on a host of outcomes as adults.

In this paper, we report the effects of a two-year cash-transfer experiment more than two years after it ended, tracking a broad range of outcomes for females aged 18-27 at follow-up.[5] Our earlier work has demonstrated the short-term effectiveness of these transfers in improving school participation and test scores, as well as reducing the incidence of pregnancy, marriage, psychological distress, and sexually transmitted infections during adolescence, indicating the possibility of finding longer-term improvements in well-being as young adults (Baird, McIntosh and Özler 2011; Baird et al. 2012; Baird, De Hoop and Özler 2013). Here, following a pre-analysis plan, we first examine human capital accumulation, marriage and fertility, labor market outcomes, and empowerment among the beneficiaries to assess the persistence of the short-term effects. Then, as most of the study participants had children at the latest follow-up, we examine their children's physical development using anthropometric measurements.

We find that the short-term improvements in the UCT arm observed during the program failed to translate into increased welfare in the longer-term. Substantial reductions in teen marriages, total live births, and HIV infections, as well as improvements in psychological wellbeing and nutritional intake observed at the end of the program were no longer apparent two years later.[6] We observe a spike in marriages and a baby boom among UCT beneficiaries immediately following the end of the program, who reported lower levels of empowerment compared with both the CCT and the control groups. However, consistent with improved physical, nutritional, and mental health during the program, we find suggestive evidence of improved height-for-age z-scores (HAZ) among children born to UCT recipients during the program.

[5] At baseline, the target population was never-married females, aged 13-22.
[6] This finding of quick convergence following short-term gains is consistent with, among others, Brudevold-Newman et al. (2017) and Hicks et al. (2018).

CCTs, on the other hand, caused sustained effects on school attainment, incidence of marriage and pregnancy, age at first birth, total number of births, and desired fertility – but only among the stratum of adolescent females who had already dropped out of school at baseline and were all assigned to CCTs. Conditional transfers were highly effective in allowing a very large share of this group to return to school.[7] However, even in this group, we find no gains in other important outcomes, such as individual earnings, per capita household consumption, subjective wellbeing, health, or empowerment. Among the stratum that was enrolled in school at baseline, while we cannot rule out some positive impacts on education and competencies, CCTs did not have any observable effects, positive or negative, on longer term outcomes of empowerment and employment. One reason behind these findings may be that the transfers were mostly inframarginal with respect to school attainment: 88% of the control group in this stratum completed primary school two years after the end of the program. Comparing the CCT and UCT groups in this stratum, we find that none of the statistically significant short-term differential impacts between the two groups remain in the longer-run – apart from age at first marriage being higher and empowerment levels lower in the UCT arm.

Our paper speaks to a small number of distinct literatures. First, it adds to a growing body of work on the medium- to long-term effects of cash transfer programs in developing countries.[8] A number of longer-term evaluations of cash transfers programs (mostly of CCTs) indicate that while cash transfer programs might improve school attainment among adolescent beneficiaries, evidence of longer-term gains in terms of learning, employment, and income are mixed as they become young adults (Baez and Camacho 2011; Behrman, Parker and Todd 2011; Barham, Macours and Maluccio 2013; Filmer and Schady 2014; Araujo, Bosch and Schady 2016; Cahyadi et al. 2018; Molina-Millán et al. 2018b). Our finding that CCT programs can substantially increase school attainment among vulnerable populations with at best mixed effects on test scores, cognitive skills, employment rates, or earnings is consistent with these studies.[9]

[7] These findings align nicely with Duflo, Dupas and Kremer (2017) who find that by age 25 Ghanaian students who were offered a secondary school scholarship were 26 percentage points more likely to complete secondary school and had 0.217 fewer children—again suggesting the importance of the magnitude of the education effect.

[8] It also builds on Baird, McIntosh and Özler (2011) and adds to the small literature that directly compares CCTs with UCTs either experimentally (Akresh et al. 2013; Benhassine et al. 2015) or quasi-experimentally (Schady and Araujo 2008; de Brauw and Hoddinott 2011; Attanasio et al. 2015).

[9] While the evidence on longer-term gains from CCT programs is mixed, especially outside of schooling (Molina-Millán 2018a) and others have also struck a cautious tone about the transformative effects of social protection programs more generally (Molyneux, Jones and Samuels 2016), some studies do find evidence of promising longer-term gains. Barrera-Osorio et al. (forthcoming) find that a forced savings treatment attached to a traditional schooling

Second, our study contributes to a large literature on the effects of programs that support pregnant women and young children. Policies for child development often target the first 1,000 days from conception to the second birthday (Barham, Macours and Maluccio 2013). What is novel in our study is that we examine the effects of targeting cash transfers to adolescent females of childbearing age and provide evidence for the important policy question on the timing of interventions for young women to protect early childhood development.[10] Our findings suggest that unconditional income support for adolescent girls and young women of childbearing age might cause significant increases in height-for-age z-scores of their children. They are also consistent with Cahyadi et al. (2018), which finds that children 0-5 who have been exposed to Indonesia's CCT program are substantially less likely to be stunted or severely stunted.[11]

The remainder of this paper is structured as follows. Section 2 describes the study setting, study design, and data collection instruments. Section 3 presents our estimation strategy. Sections 4 presents program impacts on the core respondents, followed by an examination of some key characteristics of their children. Section 5 concludes.

2. STUDY SETTING, DESIGN, AND DATA SOURCES

2.1 Study Setting

The "Schooling, Income, and Health Risk" study (SIHR) follows young women who were enrolled as never-married adolescents (aged 13-22) in Zomba, Malawi in 2007. We interviewed them for the fourth time in 2012 – approximately five years after baseline and more than two years after the cessation of the cash transfer experiment in December 2009, tracking the adolescents as

---

CCT program in Bogota increases tertiary enrollment and graduation, but the authors do not have data on other outcomes. Parker and Vogl (2018) finds that exposure to Mexico's CCT program, PROGRESA, increased school attainment by 1.3 years for both sexes, while finding statistically significant increases in labor market participation and earnings only among females. Barham, Macours and Maluccio (2018a) focus on the differential exposure of boys to Nicaragua's CCT program between the ages of 9 and 12 and find that previously demonstrated short-term increases in schooling are sustained after 10 years and there are substantial gains in learning. Barham, Macours and Maluccio (2018b) does the same for girls and finds that differential exposure to CCT does not lead to schooling gains, but it does cause reductions in fertility and increases in economic activity and earnings. The evaluation of a school-based intervention in Kenya that provided school uniforms found significant reductions in school dropout, pregnancy, and marriage among girls in the short- and medium-run; and school attainment, marriage, and childbearing by age 16 in the longer-run (Duflo, Dupas and Kremer 2015). However, another program that distributed school uniforms in Kenya found that the short-term effects on school absenteeism led to no substantive positive long-term education impacts (Evans and Ngatia 2018).

[10] Currie and Almond (2011) state "…one of the more effective ways to improve children's long-term outcomes might be to target women of child bearing age in addition to focusing on children after birth."

[11] See Manley, Gitter and Slavchevska (2013) for a review of the effects of cash transfers on children's nutritional status in low- and middle-income countries (LMICs).

many of them went on to establish their own families. These longitudinal data paint a very rich picture of the transition from adolescence into adulthood in this context. By 2012, in the control group, the study stratum that had dropped out of school at baseline had effectively completed their schooling with an average of a seventh-grade education; 81% were married, 92% had been pregnant, and only 6% had spent any time in self-employment or paid work during the past week. More than one in eight (13.5%) had been infected with HIV. The stratum of baseline schoolgirls is better-off and younger, and therefore had not proceeded as far in their transition to adulthood: in 2012, their average years of schooling was 10.4 and consistently increasing over the study period, with only 40% ever married, 50% ever pregnant, and 5.5% HIV-positive.

In the latest follow-up survey of the study sample, which was more than two years after the cessation of cash transfers, we attempted to trace the pathways through which experimentally induced changes in human capital may translate into longer-term changes in outcomes. Zomba is an almost exclusively agricultural economy characterized by low educational attainment and few opportunities for formal employment. As of 2009, this district was the third poorest in Malawi (in our sample, real monthly per-capita exchange rate comparable consumption in 2008 was USD 20.6). Secondary school completion rates are low – in our sample, among baseline schoolgirls, half of whom had completed primary school at baseline, only 17.0% had completed secondary school as of 2012. Although most adults 15 and over participate in some form of employment, the majority do not receive a formal income. In 2008, only 6% of the adult population in Zomba received a formal income (Zomba City Assembly 2009), a number that is reflected in our data with 6% of baseline dropouts and 3% of baseline schoolgirls participating in any formal work. This context is typical for many parts of rural Africa, and, hence, is an important environment in which to understand the constraints adolescents face as they transition to adulthood.

2.2 Study Design

Our study began by listing all eligible households within 176 Enumeration Areas (EAs) of the 550 EAs in Zomba District, identifying households with never-married females, aged 13-22 year-old, and dividing this target population into two main strata: those who were already out of school at baseline (*baseline dropouts*) and those who were still in school at baseline (*baseline schoolgirls*). *Baseline dropouts* comprised only 15% of target population, so were all recruited into the study. *Baseline schoolgirls* were sampled into the study at probabilities increasing in age and rural status. These two strata have always been analyzed separately for three main reasons: (a) we do not have the UCT experiment among the baseline dropouts, (b) the CCT experiment acts differently for these two groups (bringing dropouts back into school while keeping schoolgirls from dropping out) and (c) the two groups look very different based on their observable baseline characteristics. Following the pre-analysis plan, we analyze them separately in this paper as well.

Treatment was assigned first at the enumeration area (EA) level; 88 to treatment and 88 to control. All *baseline dropouts* in treatment EAs received CCTs, while we experimented with attaching conditions to the cash transfers within the larger cohort of *baseline schoolgirls*. For them, 46 EAs were assigned to CCTs, 27 were assigned to UCTs, and 15 were assigned to receive no transfers to study spillovers (from *baseline dropouts* in those EAs). The amount of money received by the household head was randomized between \$4 and \$10 at the EA level, and the core respondents were assigned their own individual transfer amounts – ranging from \$1 and \$5 – in a public lottery.[12] The share of eligible girls offered cash transfers was randomly varied across clusters to estimate spillover effects as a function of treatment intensity. Offer letters were distributed in December 2007, payments began in February 2008 and continued through the end of 2009.[13] Four rounds of data took place: Round 1-Baseline (2007), Round 2 (2008), Round 3 (2010), and Round 4 (2012). Figure I presents an illustration of the study design, and a more detailed description of the experiment can be found in (Baird, McIntosh and Özler 2011).[14]

Girls receiving UCTs simply had to show up at a local distribution point each month to pick up their transfers. Monthly school attendance for all girls in the CCT arm was checked and payment for the following month was withheld for any student whose attendance was below 80% of the number of days school was in session for the previous month. However, participants were never removed from the program for failing to meet the monthly 80% attendance rate, meaning

[12] The average total transfer to the household of \$10/month for 10 months a year is nearly 10% of the average household consumption expenditure of \$965 in Malawi in 2009 (World Bank, 2010). This falls in the range of cash transfers as a share of household consumption (or income) in other countries with similar CCT programs. The transfers were offered to all eligible girls in our target demographic and were not targeted by poverty status.

[13] In experiments like SIHR, it is important to try to understand what the beneficiaries expected as to the program's timing and duration (Bazzi, Sumarto and Suryahadi 2015). When the initial offers were made, the beneficiaries were told that the program only had funding for one year, but that efforts were being made to extend it into a two-year program. Towards the end of the first year, upon successfully obtaining additional funding, we circulated new offer letters informing the beneficiaries that the program would be continued for one more year, but not more. This message was repeated regularly at the cash distribution points by the program staff during the second and final year of the intervention.

[14] The size of the transfers, the identity of the recipients, or the intensity of treatment within the cluster did not prove to be influential on the primary outcomes of interest. Because these were randomized across the control, CCT, and UCT arms, estimates of average treatment effects remain highly robust to these controls.

that if they subsequently had satisfactory attendance, their payments would resume. Other design aspects of the program were kept identical to be able to isolate the marginal effect of imposing a schooling conditionality on outcomes of interest among *baseline schoolgirls*.[15]

2.3 Data Sources and Outcomes

The focus of this paper is data collected in Round 4, which took place in 2012, more than two years after the end of the intervention. However, to provide context for Round 4 findings, we also present impacts on the same outcomes, when available, for data collected during Rounds 2 and 3. Focusing on the core respondent, the data sources include household surveys (all rounds), biomarker data collection on HIV (Round 2-4) and Anemia (Round 4), and competencies (Round 4). In Round 4, data collection also included anthropometrics (children aged 60 months or under).

The household surveys at each round consisted of a multi-topic questionnaire administered to the households in which the core respondents resided during the data collection period. They consisted of two parts: one that was administered to the head of the household and the other administered to the core respondent. The former collected information on the household roster, dwelling characteristics, household assets and durables, shocks, and consumption. The survey administered to the core respondent collected detailed information about her family background, schooling status, health, dating patterns, sexual behavior, fertility, marriage, labor market outcomes, and empowerment.

The Round 4 household survey also included a test of basic labor market skills of the core respondent, which we termed "competencies." It included reading and following instructions to apply fertilizer; making correct change during a hypothetical market transaction; sending a text message and using a calculator on a mobile phone, and calculating profits in a hypothetical trading scenario. As Round 4 was focused more on the transition into adulthood and labor markets, as opposed to the school attainment and learning focus in Round 3, this test was designed to replace the reading comprehension, math, and cognitive skills tests utilized in Round 3, intended to serve as a measure of a more practical set of skills that might be influenced by increased schooling and needed in the labor market.

[15] For households with girls eligible to attend secondary schools at baseline, the total transfer amount was adjusted upwards by an amount equal to the average annual secondary school fees in the conditional treatment arm. This additional amount ensured that the average transfer amounts offered in the CCT and UCT arms were identical and the only difference between the two groups was the "conditionality" of the transfers on school attendance.

Home-based voluntary counseling and testing for HIV (for core respondents during Rounds 2-4) was conducted by Malawian nurses and counselors certified in conducting rapid HIV tests through the Ministry of Health HIV Unit HCT Counselor Certification Program. In addition, they tested for hemoglobin of the core respondent and measured the height and weight of all children aged 60 months or younger.

Prior to the analysis of data from Round 4, a pre-analysis plan was registered at the AEA RCT Registry (AEARCTR-0000036; https://www.socialscienceregistry.org/trials/36).[16] Our outcomes cover six pre-specified domains for the core respondent – education and competencies, marriage and fertility, health and sexual behavior, empowerment and aspirations, employment and wages, and consumption.[17] We deviate from the pre-analysis plan for the child outcomes given that the more complicated methodology needed to address causal identification was not adequately addressed in the pre-analysis plan. The analysis of program impacts on child height should thus be considered exploratory.

3. Estimation Strategy

In this section, we discuss the experimental estimation strategy used to examine program impacts on core respondents. The causal identification of program impacts on children's outcomes is more challenging and the estimation strategy used to analyze these outcomes is discussed in Section 4.6 and Appendix A.

To estimate intention-to-treat effects of the program in each treatment arm on our primary outcomes by stratum, we employ a simple reduced-form linear model:

$$Y_{ic} = \alpha + \gamma^c T_c^c + \gamma^u T_c^u + \beta X_{ic} + \varepsilon_{ic} \qquad (1)$$

where Y_{ic} is an outcome variable for core-respondent *i* in cluster *c*, T_c^C and T_c^U are binary indicators for offers in the CCT and the UCT clusters, respectively, and X_{ic} is a vector of baseline characteristics. Note that for baseline dropouts we only have the CCT binary indicator. The

[16] Many of our outcomes are in the form of indexes that are constructed using the following rubric: First, we ensured that all sub-questions are aligned so that higher scores always have a consistent meaning (good or bad). We then calculated the mean and standard deviation of the responses to each sub-question in the control group – separately for baseline schoolgirls and baseline dropouts. We then normalized each sub-question by subtracting the mean and dividing by the standard deviation. Finally we constructed (and then normalized) the raw mean of the normalized variables for all sub-questions within a family of variables to create the final index.

[17] A detailed description of all outcomes specified in the pre-analysis plan can be found here: https://drive.google.com/file/d/1hvI79ltywocFr-pafqz8_Dtg2ZXNhcHd/view.

standard errors ε_{ic} are clustered at the EA level, which account for both the design effect of our EA-level treatment and the heteroskedasticity inherent in the linear probability model.

In all regressions, we include baseline values of the following pre-specified variables as controls: a household asset index, highest grade attended, a dummy variable for having started sexual activity, and dummy variables for age in years. These variables were chosen because they are strongly predictive of schooling outcomes, hence improving the precision of the impact estimates. We also include indicators for the strata used to perform block randomization – Zomba Town, within 16 kilometers of the town, and beyond 16 kilometers (Bruhn and McKenzie 2009). Age- and stratum-specific sampling weights are used to make the results representative of the target population in the study area.

Table I presents means and standard deviations for nine individual or household characteristics for the study sample at baseline by strata and treatment assignment. As this paper is mainly about program effects more than two years after the end of cash transfers, we conduct all analysis among those who were successfully interviewed in Round 4, which maximizes sample size for the estimation of longer-term impacts.[18] Columns 1 and 2 show descriptive statistics for *baseline dropouts*, who are older than *baseline schoolgirls* and come from more disadvantaged backgrounds: for example, 44.5% of the control group had started childbearing at baseline compared to only 2.1% of *baseline schoolgirls*. In addition to the fact that all *baseline dropouts* are out of school at baseline and never married, there are no statistically significant differences between the CCT and the control groups for the variables presented in Table 1. Nor are there any differences between the two treatment groups and the control group among *baseline schoolgirls*, but the UCT group is, on average, older and has attended higher grades than the CCT group at baseline. Note that this imbalance existed at baseline and is not a result of differential attrition (Baird, McIntosh and Özler 2011). Pre-specified baseline controls used in all impact regressions described above include these two variables. Joint tests of orthogonality presented at the bottom of Table 1 confirm these findings.

[18] Conducting the analysis among the Round 4 sample implies that the Round 2 and Round 3 samples are smaller than the Round 4 sample in the analysis. For example, to be included in the Round 3 analysis of impacts, a subject had to be successfully interviewed in both Rounds 3 and 4. In addition to maximizing the sample for Round 4 analysis, which is the focus of this paper, this allows us to demonstrate that the Round 2 and Round 3 impacts, which were reported in earlier publications, hold in this sub-sample and provides some reassurance that differential attrition is not substantially affecting our findings at Round 4.

Appendix Table S1 provides tracking data for Round 4, overall and then for *baseline dropouts* and *baseline schoolgirls.* We found over 94% of the sample and surveyed close to 90% of the overall study sample split as follows across the strata and treatment arms: *baseline dropouts* lost to Round 4 follow-up: 15.7% in the control, 16.4% in treatment; *baseline schoolgirls* lost to Round 4 follow-up: 12.5% in the control, 7% in the CCT arm, and 6.7% in the UCT arm. The difference between found and surveyed is largely due to refusals, which were higher among the control group. The most common reason for refusal (which are clearly documented in the data) is no longer seeing a benefit of the study, and thus not suggestive of improved or worse outcomes for respondents who refused. Examining further the data we have on the 6% completely lost to follow-up (as opposed to found but refused to participate), general location information exists for 165 (or 86%) of the 192 remaining study participants. Coding their new location as urban if they moved to the two main cities in Malawi (Lilongwe or Blantyre) or overseas, we find no statistically significant impact of treatment on moving to urban areas in this group (in fact, the coefficient estimates are negative). While these data on our full sample partly mitigates concerns over null findings being driven by differential attrition (of successful treatment beneficiaries having moved away), we proceed with a detailed analysis of attrition.

Table II examines attrition for the same sample of core respondents who were successfully interviewed in Round 4 – first for *baseline dropouts*, then *baseline schoolgirls*. Attrition two years after the end of the cash transfer program in terms of having a completed survey is 15.7% in the control group among *baseline dropouts* and this level of attrition is not differential in the CCT arm (column 1). However, interacting attrition with the same pre-specified baseline adjustments used throughout the paper, we find that these interactions are jointly significant (column 2) – primarily because CCT beneficiaries in urban areas, which constitutes less than 20% of our sample, were more likely to be lost to follow-up[19]. Attrition in the control group among *baseline schoolgirls* is slightly lower at 12.5%, which is significantly higher than both the CCT and UCT arms (column 3). However, attrition in this stratum is not differential by baseline characteristics between treatment and control, although the F-test for joint significance of UCT interactions is 0.101 (column 4). This lack of interaction effect reflects the fact that the differential attrition is largely

[19] Note that there are only 80 respondents total in this cell, so it is unlikely to alter impact findings. Moreover, for the subset of those lost to follow-up for whom we have limited information, there is no indication of improved outcomes (they are either deceased, have a mental illness, no longer live in urban areas, or are married with no educational qualification).

driven by higher likelihood of refusals in the control group. Furthermore, and importantly for our experiment, there is no differential attrition between the CCT and UCT arms – either in levels or by characteristics. Appendix Table A2 displays the coefficients of the probit regressions that underlie the differential selection statistics provided in Table II and form the basis of the Inverse Propensity Weights (IPW) used later in the paper.

We attempt to address some of the potential bias in impact estimates due to differential attrition by treatment arm – either in levels (CCT and UCT among *baseline schoolgirls*) or in baseline characteristics (CCT among baseline dropouts) by including a thorough analysis of the robustness of our impact estimates in Section 4.5 below. There, we present estimates reweighted to account for attrition (IPW), upper and lower bounds on impact estimates for all primary outcomes (Lee 2009), as well as adjustments using the techniques of Kling and Liebman (2004). We also note that impact estimates from earlier follow-up rounds, which did not suffer from differential attrition, replicate in the Round 4 sample used in this paper.[20] These tests do rely on certain assumptions, so results should be interpreted in that light.

4. RESULTS

We start by presenting the trajectory of program effects on outcomes in four domains, separately for baseline dropouts and baseline schoolgirls: education and competencies, marriage and fertility, health, and, finally, labor market participation and empowerment.[21]

4.1 Education and Competencies

Table III presents program impacts on highest grade completed and competencies. Among *baseline dropouts*, CCTs led to an increase in highest grade completed of approximately 0.6 years, which represents a 0.22 standard deviation (SD) increase by Round 4 (Panel A). As a result, the share of beneficiaries with a Primary School Leaving Certificate (PSLC) increased by 5.8 and 8.1 percentage points in Rounds 3 and 4, respectively (Appendix Table S3, Panel A). Earlier gains in

[20] Appendix Tables S9 and S10 show the impacts on the 5 primary outcomes measured in earlier rounds using the full Round 2 and Round 3 samples when differential attrition – in levels or in baseline characteristics – was not an issue. A comparison of these estimates to those restricted to the Round 4 sample reveals no substantive differences.

[21] The reader should note that most of the one- and two-year impacts during and at the end of the program were reported in previous publications, which are clearly cited throughout the paper. What are new here are the findings from two years after the end of the program. Presenting program impacts over time within each domain allows the reader to examine the trajectory of program effects and assess whether earlier impacts were sustained. Appendix tables complement the main tables, presenting additional outcomes in all six pre-specified domains for the core respondent.

test scores of English reading comprehension, mathematics, and cognitive skills (Table III, columns 4-7) do not translate into a significant increase in scores on tests of basic labor market skills, or "competencies," such as following instructions to apply fertilizer or calculating change in a market transaction (Table III, column 8; Appendix Table S3, Panel A, columns 10-15). This result could reflect the fact that the competencies simply failed to measure variation in skills in a useful way. However, we find this explanation unlikely as the variation in schooling and test scores at the end of the intervention are strongly predictive of competencies two years later: for example, a one-year increase in highest grade completed is associated with a 0.21 SD increase in the overall competency score. Mechanically, this would imply an improvement of only 0.13 SD in the overall competency score among baseline dropouts (0.621 x 0.21 = 0.13), which is twice as large as our point estimate of 0.064 SD but well within the 95% confidence interval. Thus, we cannot rule out meaningful sustained positive impacts on skills acquisition (or zero effect) for this group.

The results for *baseline schoolgirls* suggest little, if any, effect on school attainment or competencies in either treatment group (Table III, Panel B). Any significant effect in the CCT group at the end of the program was no longer detectable two years later. The reader should note that the mean number of years completed in the control group is 10.4 in Round 4, at which point 88% of the control group had obtained a PSLC (Appendix Table S3, Panel B). Hence, while most of the transfers to *baseline schoolgirls* were inframarginal with respect to primary school completion, the cash transfer program did not cause any significant gains in secondary school completion, either. Similarly, earlier gains in test scores in the CCT group did not translate into improved competencies in the longer-run, with the only significant improvement seen in the UCT group being the ability to send a simple text message using a mobile phone. The likely explanation for these results is that the small education gains seen in the short-run lead to small (non-detectable) positive gains two years post program.[22]

4.2 Marriage and Fertility

[22] We chose not to pre-specify self-reported enrollment as a primary outcome based on findings from Baird, McIntosh and Özler (2011) that there is significant (and differential) misreporting in this variable. With this caveat in mind, we do still find that core respondents in the CCT and UCT arm report enrollment rates that are approximately six percentage points higher ($p<0.05$) than the control at Round 4. Baseline dropouts are also four percentage points more likely to be enrolled in school ($p<0.01$), over a base of 2.4%.

As with the education outcomes, CCTs had large effects on marriage and fertility for *baseline dropouts* that were sustained at Round 4 (Table IV, Panel A). They were 14.0, 15.7, and 10.7 percentage points (pp) less likely to have been ever married at Rounds 2-4, respectively (all significant at 99% confidence). The corresponding reductions were 5.7, 8.1, and 4.0 pp for being ever pregnant (all significant at 90% confidence or higher). Furthermore, there is a negative fertility gradient among CCT beneficiaries, leading to a reduction of 0.147 total live births at Round 4 (p-value < 0.001), which corresponds to a reduction of more than 10% (0.19 SD) and is consistent with the reduction in stated desired fertility. Age at first marriage and first birth, both calculated at the intensive margin, were similarly higher by 0.43 and 0.27 years, respectively.

Among *baseline schoolgirls*, CCTs had no effects on marriage and fertility at any point during our study period (Table IV, Panel B). On the other hand, UCTs were very effective in substantially reducing marriage and pregnancy rates among baseline schoolgirls during and at the end of the program (Baird, McIntosh and Özler 2011). Two years later, there are no longer any differences in ever married, ever pregnant, total number of live births, or even age at first birth between the UCT group and either the control group or the CCT arm. We find that the age at first marriage is delayed by half a year as of Round 4, which is consistent with the fact that girls in the UCT arm who delayed marriage got quickly married following the end of the intervention. Striking spikes in pregnancies and marriages in the UCT group immediately following the end of the transfers are shown in Figure II. The temporary nature of the fertility changes in this group is also reinforced by the fact that desired fertility remains unchanged (Table IV, Panel B, column 12).[23] In analysis not shown in the tables, we find that teen pregnancy (defined as starting childbearing at age 18 or younger) was significantly lower in the UCT arm at Round 3 (3.8 pp, p-value =0.027) but that this effect had also shrunk by two thirds and was no longer significant by Round 4. Beneficiaries of all ages experienced a spike in marriages and pregnancies following the program, meaning that UCTs reduced the prevalence of neither teen pregnancies nor child marriages by Round 4 – despite large reductions in these quantities at Round 3.

[23] The finding of null effects in Round 4 in the UCT arm is not simply a function of lack of power. While the standard errors of binary indicators for marriage and pregnancy are higher in Round 4 than in Round 3 due to the fact that the control means for these variables are increasing towards 0.5 over the course of our study, minimum detectable effects as a percentage of the mean in the control group are actually lower. Furthermore, these minimum detectable effects are comparable to or lower than those presented in similar papers, such as Bandiera et al. (2017b). Finally, many of the significant effects among *baseline dropouts* that we present in Table IV are larger than the minimum detectable effects among *baseline schoolgirls*.

Cash transfers can have effects on marriage and fertility via two channels. The first pathway, apparent in the UCT arm, is through an income effect. In our study, this effect is strong but disappears immediately when the transfers stop – as the transfers have not led to any accumulation of physical or human capital. The other pathway, apparent in the CCT arm among *baseline dropouts*, is through increased schooling. Increased schooling is strongly associated with delays in marriage and childbearing and reductions in desired and total fertility, but the impacts of transfer programs on schooling need to be substantial to translate into meaningful and statistically significant knock-on effects on marriage and fertility.

4.3 Health

Table V presents program impacts on biomarkers for HIV and anemia – the primary health outcomes specified in our pre-analysis plan. Program effects on HIV prevalence during the program, i.e. at Round 2, were reported in Baird et al. (2012). Despite the improvements in education, delays in marriage and fertility, and the high prevalence of HIV among *baseline dropouts* (13.5% by Round 4), CCTs did not reduce HIV prevalence in this stratum at any point during the study period (Panel A). Appendix Tables S4 and S5 examine self-reported sexual behavior on the extensive and intensive margin. Both the onset of sexual activity and the likelihood of being sexually active during the past year were lower among program beneficiaries during and immediately after the program, but not two years later. There were no effects on risky sexual behavior, such as having older partners or use of condoms, among those who reported being sexually active. Nor did CCTs have significant effects on psychological wellbeing or nutritional intake (Appendix Table S6).

Among *baseline schoolgirls*, program impacts on HIV mirror those on marriage and fertility over time: there is no effect of CCTs on HIV at Rounds 3 or 4, but a more than 50% reduction in HIV prevalence in the UCT group (albeit, statistically significant only at the 90% level) at the end of the intervention is no longer there two years later (Table V, Panel B). During the two-year post-intervention period, which saw a spike in pregnancies and marriage in the UCT group, the incidence of HIV was 3.5 percentage points (pp) – compared with 2.0 pp in the control group, though not statistically significant. Appendix Table S6 shows that effects of cash transfers were equally transient on mental health and nutritional intake – strongly evident during the program and disappearing afterwards. There is weak evidence of lower anemia prevalence in the

UCT arm in Round 4, but this finding is not robust to either using a continuous measure of hemoglobin levels, or to multiple hypothesis testing corrections presented in Section 4.5.

4.4 Labor Market Participation and Empowerment

The main activities performed by the *baseline dropouts* in our sample are household chores – such as cooking and cleaning, fetching water and firewood, and looking after children – (69.6%) and subsistence agriculture (19.4%); among *baseline schoolgirls*, 55.2% report household chores as their main activity, 11.1% report subsistence agriculture, while 27.5% are still in school. Hardly anyone in our sample spent a significant amount of time in self-employment or paid work during the past week (Table VI column 3), consistent with other data on labor market participation in Zomba. Only a third of baseline dropouts and a quarter of baseline schoolgirls report having done any wage work in the past three months (Appendix Table S7). There are no significant effects on primary outcomes in either stratum, except a negative effect on typical wage among *baseline dropouts*, which may reflect the fact that individuals in the treatment group were in school longer, and thus might have less work experience. Program impacts on secondary labor market outcomes, such as the effective daily wage, labor income in the past five seasons, and any wage work in the past three months, are similarly null (Appendix Table S7).[24]

When we conduct exploratory analysis to investigate broader questions of time use, we find that *baseline dropouts* in the treatment group are still spending more hours in school (1.54 hours per week, p=0.018), which may explain the negative effect on typical wage. When we look at proportion of hours in school or work the impact is insignificant and zero (-0.001, p=0.930), indicating that these additional hours in school are completely offset by additional hours in work by the control group. For *baseline schoolgirls*, we find that respondents in the CCT arm are spending approximately 3.63 hours per week more in school (p=0.045) and that this does translate to more time in work and school (2.4 percentage point more time in work or school, p=0.063). For UCT recipients there are no impacts on either additional hours in school (-0.088, p=0.964) or time in work and school (0.001, p=0.919). This result supports the possibility of small positive sustained impacts on time allocation for *baseline schoolgirls* in the CCT arm, with clear null effects in the UCT arm.

---

[24] We also examined accumulation of savings, household assets, and productive assets (such as livestock). We find no treatment effects on any of these outcomes in either stratum.

For *baseline dropouts,* program impacts on empowerment echo those on competencies, health, and labor market participation: despite significant gains in educational attainment, delays in marriage and pregnancy, and reductions in total live births, there are no effects on the overall index of empowerment or subjective welfare (Table VI, Panel A, columns 4 & 5) and, in fact, almost all coefficient estimates are negative. This finding holds when we examine empowerment by marital status, i.e. on the intensive margin, at Round 4 (columns 6 & 7). Appendix Table S8 shows estimates for the components of the female empowerment index (self-esteem, social participation, preferences for child education, and aspirations).

For *baseline schoolgirls* in the CCT group, we also see no significant impacts on empowerment or subjective wellbeing, although the coefficient estimates are generally positive. However, in the UCT arm, the empowerment index is significantly lower than both the control and the CCT groups (Table VI, Panel B). The -0.159 SD effect (p-value=0.05) on the super-index of overall empowerment among the UCT beneficiaries is reflected in the negative (but insignificant) effects in all sub-indices except aspirations (Appendix Table S8, Panel B), and is driven mainly by a large (-0.342 SD; p-value<0.01) and significant negative association with empowerment among those who are married (Table VI, Panel B, column 7). The findings indicate a statistically significant divergence in female empowerment between CCT and UCT recipients among baseline schoolgirls two years after the end of the cash transfer program – particularly for those married by Round 4, which may be related to the spike in marriages immediately after the cessation of cash transfers (Figure II, Panel B).

4.5 Robustness of Findings to Attrition and Multiple Hypothesis Testing

Before we move on to analyzing child outcomes, we examine the robustness of program impacts for the young women targeted by our cash transfer program. We address two concerns with our primary analysis: attrition and multiple hypothesis testing. First, in Section 3, we showed that while the share of our study sample lost to follow-up more than four years after baseline data collection is not high (between 12.5% and 15.7% in the control groups of the two strata), there is evidence of differential attrition in levels (but not characteristics) among *baseline schoolgirls*, and vice versa among *baseline dropouts*.[25] We reiterate, however, that there is no differential attrition

[25] In addition, as mentioned above, a comparison of impact estimates in Rounds 2 & 3 using all the available data for each of those rounds vs. the samples restricted to those available in Round 4 show qualitatively the same impacts.

in levels or characteristics between the CCT and UCT arms among *baseline schoolgirls.* As differential attrition has the potential to bias impact estimates and, as such, is a threat to causal inference, we conduct additional analysis to test the robustness of our findings. Second, as we present impacts on 14 pre-specified primary outcomes in Round 4, we present *q-values* for impact estimates that are adjusted for the false discovery rate (FDR) – to allay concerns that some of the statistically significant impacts estimates might have materialized by chance.

In Appendix Tables S11-S13, we examine the potential effects that differential attrition may have had on our results. In these tables, we present a central column (4) that re-states the impacts shown earlier in the paper, using only sampling weights. In column (5), we estimate and implement attrition propensity weights, first running a probit regression predicting presence in the Round 4 sample with our standard battery of baseline covariates and their interaction with treatment, and then weight outcomes by the product of the sampling weights and the inverse of this follow-up success probability. In columns (3) and (6) we present, respectively, the lower and upper bound estimates trimming the tails of the distribution following Lee (2009) to generate the same observed attrition rates both treatment arms. In the remaining columns we follow Kling and Liebman (2004) and impute to the missing observations the mean within that treatment arm plus or minus 0.1 * the arm-specific standard deviation (columns (2) and (7)) and plus or minus 0.25 * the standard deviation (columns (1) and (8)). For the lower bounds this amount is subtracted from the treatment and added to the control, and for the upper bounds this is reversed.

For *baseline dropouts*, we note that the Lee bounds are tight around the original estimate because the difference in the level of attrition between the control and the CCT groups is very small (Appendix Table S11). Furthermore, IPW-adjusted impact estimates are very close to our original estimates. Even the Kling and Liebman bounds present a very consistent picture of impacts; across the full set of bounds we find significant impacts of the CCT on highest grade completed, ever married, and number of live births. Nothing in the table suggests that we should significantly revise our interpretation of the key findings of program impacts among *baseline dropouts*. Similarly, for *baseline schoolgirls*, IPW-adjusted estimates are nearly indistinguishable from the original estimates, while the Lee bounds are wider because of the larger difference in attrition levels between the control group and either treatment group (Appendix Table S12). For *baseline schoolgirls* there are no outcomes that are significant across bounding strategies, although the negative effects of unconditional treatment on empowerment and anemia come close. These wider bounds mean that while our original and IPW-adjusted estimates generally indicate a lack of impact of CCTs or UCTs among *baseline schoolgirls* in Round 4, we cannot rule out sizeable (positive) impacts for some of the more intermediate outcomes, such as highest grade completed and competencies.

Finally, Appendix Table S13 shows that pairwise comparisons of CCT and UCT impacts are completely robust to the adjustments we implement, which confirm that (a) most of the statistically significant differences in schooling, marriage, and fertility that existed between these two treatment arms immediately after the program disappeared two years later, and (b) UCT beneficiaries have a lower level of overall empowerment than CCT beneficiaries by Round 4.

In summary, the analysis above confirms that our medium-term impact findings among *baseline dropouts* and the comparison of relative impacts of CCTs vs. UCTs among *baseline schoolgirls* are strongly robust to alternative means of handling attrition in the data. On the other hand, for comparisons of CCTs or UCTs with the pure control group among *baseline schoolgirls*, it is prudent to allow for the possibility of positive impacts on education and health and negative effects on labor market participation (perhaps due to higher likelihood of being in school).

In Table VII, we present *q-values* controlling for FDR, as described in Anderson (2008). We use Anderson's Stata code to calculate FDR-adjusted *q-values*, which uses a simple method proposed by Benjamini and Hochberg (1995) to calculate the smallest *q* at which each hypothesis would be rejected.[26] The *q-values* for the 14 primary outcomes in this study are calculated separately for each key comparison: CCT vs. Control among *baseline dropouts*; and then CCT vs. Control, UCT vs. Control, and CCT vs. UCT among *baseline schoolgirls*. These estimates, presented alongside the original *p-values* of the impact estimates for each treatment arm, confirm the robustness of our findings to multiple hypothesis testing adjustments: every statistically significant impact for the CCT arm among *baseline dropouts* has a *q-value* below 0.099, while every *q-value* is greater than 0.289 among *baseline schoolgirls*.

Our analysis so far points to two main findings: first, among the more vulnerable group of baseline dropouts, CCTs improved school attainment and decreased marriage and fertility rates, which were sustained over time. Second, the large effects of UCTs among *baseline schoolgirls*

[26] The Stata code and the paper that describes the method can be found here: https://are.berkeley.edu/~mlanderson/ARE_Website/Research.html.

during the program have all but disappeared within two years. In this sub-section, we found that these two main findings are robust to attrition and multiple hypothesis testing.

4.6 Child Outcomes

We conclude this section with a discussion of program impacts on children born to study participants. Policies for child development often target the first 1,000 days – from conception to the second birthday (Barham, Macours and Maluccio 2013), a period during which improvements in family income may be particularly important for children's development.[27] In our experiment, more than 2,000 babies were born to study participants by Round 4 – with endogenous variation in their duration of exposure to the cash transfer program. We have already demonstrated that well-known channels for growth, such as maternal nutrition and stress (Black, Devereux and Salvanes 2016), improved during the two-year program.

In terms of the timing and structure of the cash transfers, we would expect substantial heterogeneity of program impacts on child outcomes both by when the birth took place and whether the transfers to the mother were conditional on school attendance. As in other countries in the region, childbearing and schooling are mutually exclusive in Malawi (Baird, McIntosh and Özler 2011; Ozier 2015), meaning that the condition to regularly attend school effectively screens out most expecting and new mothers in the CCT arm: only in the UCT arm would mothers with newborn children continue receiving transfers. Secondly, even in the UCT arm, a child conceived after the end of the program would have had no direct exposure to the program and, as we have shown earlier, the average mother would have acquired no additional education that could provide subsequent human capital-driven benefits. On the other hand, increased mother's education can, for example, increase child height (Thomas, Strauss and Henriques 1991), so we might expect to see benefits among children born after the program in the CCT groups – particularly among baseline dropouts, who experienced large gains in school attainment themselves. These causal chains suggest that UCT benefits should be concentrated among children born or in utero during

[27] Agüero, Carter and Woolard (2006) study the effect of Child Support Grants in South Africa for children who were exposed to the program up to three years after birth and find sizeable effects of increased exposure to these unconditional cash transfers on child height. Milligan and Stabile (2009), studying child benefits in Canada, find effects on cognitive and socio-emotional skills of children aged 4-6. Dahl and Lochner (2012) using the variation in Earned Income Tax Credit in the U.S., find that increased income improves children's test scores. Currie and Almond (2011) review the effects of "near cash" programs, such as food stamps, in the U.S. and find credible evidence of effects on birth weight. Finally, Aizer et al. (2016) and Hoynes, Schanzenbach and Almond (2016) find that children whose parents received cash transfers and food stamps in the U.S. had improved education, health, and income as adults.

the program, while CCTs might be most beneficial to children born after the mother's additional human capital accumulation took place.[28]

Since the program caused significant changes in fertility patterns (Table IV), the raw treatment-control differences in, say, height are not interpretable as causal impacts of the program on a specific child, because childbearing is endogenous to treatment. To address this, we pursue two approaches: estimation of heterogeneity by child age, and regression/reweighting control for selection-driven covariates. The technical details of the assumptions that are required for this approach and the sequence of adjustments that we made are outlined in Appendix A. We concentrate our analysis on height-for-age z-scores (HAZ), which is an objectively measured indicator of stunting that affects almost 50% of children under the age of five in Malawi, and is a strong predictor of productivity as an adult in low income settings (LaFave and Thomas 2016).[29]

First, we can examine how treatment effects vary across three 'epochs' defined by child age. The first epoch captures those *directly exposed* to the program, meaning those born during the program.[30] The second epoch covers those born within nine months of the end of the program, who were *exposed in utero* for a maximum of nine months. Finally, the third epoch covers those born more than nine months after the end of the program, who were *not exposed* to cash transfers either as children or in utero and could only benefit from the program due to improved outcomes of their mothers**.**

Figure III plots the "raw" differences in HAZ for children under 60 months between the treatment and the control groups (the thinner curves in plain font).[31] The figures are consistent with the hypothesis that differences in children's heights are moderated by exposure to the

[28] Increased age at first birth can also have positive effects on child height through improved gynecological maturity and decreased competition for nutrition between the mother and the child in utero, which could operate in any treatment group that delayed pregnancies.

[29] Of the two anthropometric measures that we collected for children aged 0-59 months – height and weight – stunting (height-for-age z-score<-2) is the key indicator of malnutrition in Malawi: almost half of the children under the age of 5 were categorized as stunted in 2010, while wasting (weight-for-height z-score<-2) rates are low at 4% (IFPRI 2014). Child assessments are also objectively measured outcomes of cognitive and socio-emotional development, but the target age group for the assessments that we chose for this study (36-59 months) makes them unsuitable for analysis by epoch of exposure to the program because only children born during the first year of the program (less than 200 in the baseline schoolgirl stratum with less than 30 in the UCT arm) were eligible for assessment.

[30] The percentage of baseline schoolgirls who reported having been ever pregnant was less than 2% at baseline. Hence, children directly exposed to the program in this stratum are almost exclusively born during the intervention. However, approximately 45% of baseline dropouts had already started childbearing at baseline. Therefore, our analysis includes children under two at the start of the program, who were at least partially exposed to cash transfers.

[31] We construct these figures by running a locally weighted treatment effects regression across the distribution of child age (Fan 1992) and plotting the resulting time-specific treatment effects and 95% confidence intervals. The figures plot robust standard errors clustered at the EA level, as in the main regressions.

program. Most strikingly, we see a very large difference in HAZ between the UCT and the control group during the program, which steadily declines, disappears by the end of the program, and even turns negative during the final epoch (Panel C). This pattern is consistent with the substantive but transient improvements in the nutritional status and mental health of UCT beneficiaries. In contrast, no significant differences in child height are apparent between the CCT and the control groups during the program – also consistent with the fact that most mothers of children born in this period would have dropped out of school because of their pregnancies, thus forgoing any cash transfers (Panels A and B). Column 1 in Tables VIII and IX reports the raw differences in HAZ by epoch, for baseline dropouts and baseline schoolgirls respectively, and confirms these patterns.

Next, we attempt to control for endogenous selection via propensity weighting and regression control. The treatment/control comparisons may combine extensive margin selection effects (such as the types of women who became pregnant, the types of partners they chose, and the age at birth) with a 'direct' casual effect of the program on the identities of the children observed in the Round 4 sample. Unlike many such applications in the natural experimental literature, it is entirely plausible that *all* of the observed impacts on HAZ arise from the selection effect of unwanted children being delayed by the receipt of the UCT.[32] Following the methodology laid out in Appendix A, we can then sequentially implement a set of selection controls: in Column 2 we use a set of baseline maternal characteristics to predict fertility in each epoch, and include inverse propensity weights based on fertility probabilities in the analysis (as well as including these covariates in the regression) to provide estimates of impact that are doubly robust to maternal type selection. Column 3 includes covariates controlling for paternal type, Column 4 adds flexible controls for child age, while Column 5 adds indicator variables for the mother's age at birth and interactions of maternal age with all other baseline covariates. Subject to the assumptions laid out in Appendix A, these estimates allow us to move from the reduced-form 'raw' treatment effects to estimates of a 'direct' effect – i.e. suggestive *ceteris paribus* impacts of CCTs and UCTs on the children born in each epoch.

[32] In the study of a negative shock, the most likely extensive margin impact is an increase in mortality among the weakest fetuses and children, hence pushing upwards the average outcome among surviving cohorts exposed to the shock. The large set of papers studying negative shocks such as pollution (Chay and Greenstone 2003; Adhvaryu et al. 2016; Black et al. 2017), disease (Almond 2006), and hunger (Almond and Mazumder 2011) can typically argue that any negative effects found on surviving children are conservative. Because we study a positive shock that may have delayed pregnancies with worse expected outcomes, the selection and direct treatment effects in our case may both point to superior child outcomes in the treatment condition. Decomposing these effects is therefore critical.

Column 2 in Table IX, Panel A shows that the maternal selection controls alone reduce the effect of UCTs during the program by almost a half (from .953 to .525 SD), confirming significant positive selection into childbearing during the program in the UCT arm. The other pathways have a limited effect, resulting in a fully adjusted direct effect of .523 SD (column 5). The size of this remaining direct effect is consistent with Barham, Macours and Maluccio (2013), who report that children in Nicaragua who received three years of cash transfers were 0.2-0.4 SD taller; and with Agüero, Carter and Woolard (2006), who find that children in South Africa receiving child support grants for most of the period between 0-3 years of age gained as much as 0.45 SD in HAZ.[33] The bold curves in Figure III plot these 'direct', fully adjusted Fan regressions across the month of birth, including the battery of controls included in Column 5 of Tables VIII & IX. The distribution of direct treatment effects in the UCT arm shown in Panel C is remarkably consistent with what we would expect: a significant and positive effect on HAZ among children born during the program, which disappears immediately following the cessation of transfers.[34]

The effects on HAZ in the CCT groups are also as expected: as females who dropped out of school due to pregnancies did not continue to receive transfers, we would expect little effect on their children born during the program. Conversely, if increased education or delaying childbearing influences child height, we might see effects among children of CCT recipients born after the program. Among *baseline dropouts* or *baseline schoolgirls*, we see no significant effects on HAZ for babies born during the program. However, the corrected tables show modest (0.10-0.25 SD), but statistically non-significant, improvements in HAZ for children born after the program to *baseline schoolgirls* who received CCTs (Table IX, Panels A & C, column 5).

The findings here are consistent with the theory that underlies the tradeoff between CCTs for schooling and UCTs: UCTs primarily confer an income effect on children born during the program and no effects on children born later because they do not lead to an accumulation of capital (human, physical, or social) for the mother.[35] On the other hand, CCTs deny such benefits

---

[33] Examining an ongoing CCT program in Indonesia, Cahyadi et al. (2018) find reductions of 23 to 27 percent in the probability of being stunted (and 56 to 62 percent in the probability of being severely stunted) among children aged 0-5 six years after the start of the program.

[34] If there is a pure income effect on child height, it is possible that this effect responds to increased transfer amounts. In Appendix Table S14, we investigate the effect of randomly assigned transfers to the core respondent and her household separately. These estimates provide suggestive evidence of decreased psychological distress and increased consumption of meals with animal proteins as a function of unconditional transfer amounts to the core respondent.

[35] We do not see any positive effects of UCTs for babies born within nine months from the end of the program, i.e. those *exposed in utero*. While this may be considered surprising given the extant evidence on the importance of this period for physical development, it should be remembered that the young mothers are also dealing with the cessation

to the children of non-compliers during the program, but may have modest effects on future children through increased human capital accumulation.

5. CONCLUSION

The most striking feature of the findings presented in this paper is the transience of the impact of unconditional cash transfers. Particularly glaring are the fleeting decreases in child marriage and teen pregnancy in the UCT arm, along with psychological distress and HIV – the prevalence of all of which reverted to control group levels within just two years, implying significant but temporary income effects. Within months of the end of the program, many UCT beneficiaries became pregnant, and were married soon thereafter.

On the other hand, there were sustained program effects on school attainment, early marriage, and pregnancy for *baseline dropouts* receiving CCTs. However, these effects did not translate into reductions in HIV, gains in labor market outcomes, or increased empowerment.[36] Several reasons might explain the disconnect between increased school attainment and no improvements in labor market outcomes, empowerment, or health. First, it is possible that increased schooling does not provide one with the skills needed to increase future welfare in this context. There are very few formal sector jobs for women in Malawi and most households depend on subsistence farming and a variety of informal sector activities. We administered tests of skills needed in farming and running small household enterprises and detected no effects in these domains. If safe and well-paying jobs existed for women in Malawi, households might invest in the necessary human capital of adolescent females on their own – perhaps even without the help of any outside interventions (Munshi and Rosenzweig 2006; Jensen 2012; Oster and Steinberg 2013; Heath and Mobarak 2015). Second, task performance is dependent on not only improvements in cognitive skills, but also on character skills and effort (Heckman and Kautz 2013). Hence, it is possible that CCTs, by providing incentives for formal schooling, improved only cognitive skills, which may not have been enough to increase productivity.[37]

---

of support during this same period. Changes in lifestyle and increased stress from the loss of regular income during this transitional period may have dampened any beneficial effects of cash transfers on the child *in utero*.

[36] The findings on marriage, pregnancy, and HIV are consistent with Duflo, Dupas and Kremer (2015), who find that education subsidies in Kenya reduce dropout, pregnancy, and marriage, but not sexually transmitted infections. They suggest a model in which choices between committed and casual relationships, rather than unprotected sex alone, affect pregnancy and HIV.

[37] Heckman and Mosso (2014) state "The most effective adolescent interventions target formation of personality, socioemotional, and character skills through mentoring and guidance, including providing information." Bandiera et

Our study provides some important guideposts for the design of effective adolescent-focused cash transfer programs. First, the palliative benefits of small and frequent unconditional cash transfers are uncontested and reinforced by our study, but the idea that they can contribute to a sustained improvement in welfare over the longer-run is unproven and not supported here.[38] Second, we shed further light on the tradeoffs between the benefits of conditional and unconditional transfers. The lack of knock-on effects from schooling gains in this context implies that the imperative to use conditions to generate increased investments in human capital may be weak when few income-generating opportunities exist. Moreover, by denying noncompliant adolescent girls and young women cash transfers at precisely the moment when they are most likely to start childbearing, a myriad of potential benefits is missed under CCT programs.

A potentially promising way of resolving this tradeoff is to view CCT and UCT programs as complements to each other rather than alternatives: policymakers could provide a basic unconditional cash transfer to adolescent girls topped up by conditional cash transfers for human capital accumulation and desired health behaviors – providing both an incentive to invest in education and health while still guaranteeing a basic level of protection to those who are unable or unwilling to comply with the conditions. Third, and finally, the promising (if only suggestive) evidence of the positive effect of UCTs on children's height provides an additional reason to consider providing basic UCTs to adolescent females. Indeed, Currie and Almond (2011) have suggested that targeting transfers towards women of childbearing age may be beneficial in the U.S. context, so as to maximize benefits to children *in utero*. This form of targeting would suffer from remarkably little 'leakage' in the Malawian context; two thirds of women aged 20-24 gave birth by age 20 and virtually all females have started childbearing by age 25 (National Statistical Office and ICF Macro 2005).

Our study has some limitations. First, differential attrition levels between either treatment arm and the control group among *baseline schoolgirls* reduces the precision of and the confidence in the null impact estimates in Round 4. Second, we examine the height of children born to study participants, which itself is endogenous to treatment. Therefore, the non-experimental impact

---

al. (2017b) provide suggestive evidence that a mentoring program in Uganda (ELA) that provided young females with "hard" vocational and "soft" life skills may have led to longer-term improvements in welfare.

[38] We do not mean to downplay or underestimate the effects of redistributive policies on current poverty and inequality reduction, even if they do not lead to substantive increases in human capital accumulation among adolescents. Welfare gains from such effects can be as large as, if not larger than, those from human capital investments (Alderman, Behrman and Tasneem 2015).

estimates on HAZ should be treated as suggestive. Finally, the study sample of initially never-married females, aged 13-22 had not yet completed their transition from adolescence to adulthood by Round 4, especially in the *baseline schoolgirl* stratum. Hence, it is important to allow for the possibility that clearer program impacts may emerge in a future round of data collection on important outcomes such as skills, labor market participation, earnings, and empowerment.

Given the medium-term nature of these results, it is natural to ask how much we can infer about longer-run impacts. As our study captures outcomes a little more than two years after the cash transfers stopped, we cannot speak to long-term effects, such as those analyzed in the U.S. context in recent studies (Aizer et al. 2016; Hoynes, Schanzenbach and Almond 2016). To guide our thinking, we return again to the role of productive assets in generating long-term rewards: to make an impact later in life, a program must have meaningfully shifted the stock of some form of capital that can generate returns over the long haul. For *baseline dropouts*, who were offered CCTs to return to school, the increase in school attainment, and the subsequent drop in fertility, is sizeable. For this group, it may be premature to conclude that improvements in education will lead to no long-term gains. If the relationship between education and wages becomes steeper with age, or if household-level human capital alters the economic trajectory of these households, future follow-up studies may well reveal longer-term benefits. For *baseline schoolgirls* in the UCT arm, our findings suggest that two years of financial support during adolescence might have been too short – rather than a two-year follow-up window being too short to trace out subsequent impacts.[39] Only two years after the end of the program, UCT beneficiaries are, in most respects, in a position indistinguishable from where they would have been in the absence of cash transfers. The unwinding of the program impacts on marriage and pregnancy is immediate and substantial, so, given the lack of school attainment or learning effects in this group, it is only their children in whom we note some vehicle for durable improvements in human capital.

[39] However, it should be noted that the Mothers' Pension program of the early 20th century U.S. had a median duration of three years and was of similar generosity to many cash transfer programs today (Aizer et al. 2016), including ours, and showed long-term effects in health, education, and income among children of program beneficiaries.

REFERENCES

Adhvaryu, Achyuta, Prashant Bharadwaj, James Fenske, Anant Nyshadham, and Richard Stanley. 2016. Dust and Death: Evidence from the West African Harmattan. *CSAE Working Paper WPS/2016-03.* Centre for the Study of African Economies, University of Oxford.

Agüero, Jorge, Michael Carter, and Ingrid Woolard. 2006. The Impact of Unconditional Cash Transfers on Nutrition: The South African Child Support Grant. SALDRU Working Paper Series No. 06/08.

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney. 2016. The Long-Run Impact of Cash Transfers to Poor Families. *American Economic Review* 106 (4): 935-971.

Akresh, Richard, Damien de Walque, and Harounan Kazianga. 2013. Cash Transfers and Child Schooling: Evidence from a Randomized Evaluation of the Role of Conditionality. *Policy Research Working Paper No. 6340*. World Bank, Washington, DC.

Alderman, Harold, Jere Behrman, and Afia Tasneem. 2015. The contribution of increased equity to the estimated social benefits from a transfer program: An illustration from PROGRESA. IFPRI Discussion Paper 1475.

Almond, D. 2006. Is the 1918 Influenza Pandemic Over? Long-Term Effects of *In Utero* Influenza Exposure in the Post-1940 U.S. Population. *Journal of Political Economy* 114 (4): 672-712.

Almond, Douglas and Bhashkar A Mazumder. 2011. Health Capital and the Prenatal Environment: The Effect of Ramadan Observance during Pregnancy. *American Economic Journal: Applied Economics* 3 (4): 56-85.

Anderson, Michael L. 2008. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association* 103 (484): 1481-1495.

Araujo, M Caridad, Mariano Bosch, and Norbert Schady. 2016. Can Cash Transfers Help Households Escape an Inter-Generational Poverty Trap? *NBER Working Paper 22670.* National Bureau of Economic Research.

Attanasio, Orazio P., Veruska Oppedisano, and Marcos Vera-Hernández. 2015. Should Cash Transfers Be Conditional? Conditionality, Preventive Care, and Health Outcomes. *American Economic Journal: Applied Economics*, 7(2): 35-52.

Baez, Javier Eduardo and Adriana Camacho. 2011. Assessing the long-term effects of conditional cash transfers on human capital: evidence from Colombia. *IZA Discussion Paper No. 5751*.

Baird, Sarah, Craig McIntosh, and Berk Özler. 2011. Cash or Condition? Evidence from a Cash Transfer Experiment. *Quarterly Journal of Economics* 126 (4): 1709-1753.

Baird, Sarah, Craig McIntosh, and Berk Özler. 2016. When the Money Runs Out: Do Cash Transfers Have Sustained Effects on Human Capital Accumulation? *World Bank Policy Research Working Paper*, no. 7901.

Baird, Sarah, Francisco HG Ferreira, Berk Özler, and Michael Woolcock. 2013. Relative effectiveness of conditional and unconditional cash transfers for schooling outcomes in developing countries: a systematic review. *Campbell Systematic Reviews* 9 (8).

Baird, Sarah, Jacobus De Hoop, and Berk Özler. 2013. Income Shocks and Adolescent Mental Health. *Journal of Human Resources* 48 (2): 370-403.

Baird, Sarah and Berk Özler. 2016. Sustained Effects on Economic Empowerment of Interventions for Adolescent Girls: Existing Evidence and Knowledge Gaps. CGD Background Paper: http://www.cgdev.org/sites/default/files/sustained-effects-economic-empowerment.pdf.

Baird, Sarah J, Richard S Garfein, Craig T McIntosh, and Berk Özler. 2012. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. *The Lancet* 379 (9823): 1320-1329.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017a. Labor Markets and Poverty in Village Economies. *The Quarterly Journal of Economics* 132 (2): 811-870.

Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017b. Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa. *Unpublished manuscript*:

http://www.ucl.ac.uk/~uctpimr/research/ELA.pdf.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science* 348 (6236): 1260799.

Barham, Tania, Karen Macours, and John A Maluccio. 2013. Boys' Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages for Early Childhood Interventions. *American Economic Review* 103 (3): 467-471.

Barham, Tania and Karen Macours, and John A Maluccio. 2017. Are Conditional Cash Transfers Fulfilling Their Promise? Schooling, Learning, and Earnings after 10 Years. CEPR Discussion Paper No. DP11937.

Barrera-Osorio, Felipe, Leigh L Linden, and Juan E Saavedra. 2017. Medium- and Long-Term Educational Consequences of Alternative Conditional Cash Transfer Designs: Experimental Evidence from Colombia. *NBER Working Paper 23275*. National Bureau of Economic Research.

Bazzi, Samuel, Sudarno Sumarto, and Asep Suryahadi. 2015. It's all in the timing: Cash transfers and consumption smoothing in a developing country. *Journal of Economic Behavior & Organization* 119:267-288.

Behrman, Jere R, Susan W Parker, and Petra E Todd. 2011. Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? *Journal of Human Resources* 46 (1): 93-122.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas and Victor Pouliquen. 2015. "Turning a Shove into a Nudge? A "Labeled Cash Transfer" for Education." *American Economic Journal: Economic Policy*, 7(3): 86-125.

Benjamini, Yoav and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 289-300.

Bianchi, Milo and Matteo Bobba. 2013. Liquidity, risk, and occupational choices. *The Review of Economic Studies* 80 (2): 491-511.

Black, Sandra E, Aline Bütikofer, P Devereux, and K Salvanes. 2017. This Is Only a Test? Long-Run and Intergenerational Impacts of Prenatal Exposure to Radioactive Fallout. *NBER Working Paper 18987*. National Bureau of Economic Research.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2016. Does Grief Transfer across Generations? Bereavements during Pregnancy and Child Outcomes. *American Economic Journal: Applied Economics* 8 (1): 193-223.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2014. Generating skilled self-employment in developing countries: Experimental evidence from Uganda. *Quarterly Journal of Economics*, 129 (2): 697-752.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2018. The Long-Term Impacts of Grants on Poverty: 9-Year Evidence from Uganda's Youth Opportunities Program. *Unpublished Manuscript*.

Brudevold-Newman, Andrew, Maddalena Honorati, Pamela Jakiela, and Owen Ozier. 2017. A Firm of One's Own: Experimental Evidence on Credit Constraints and Occupational Choice. *Unpublished Manuscript.* https://sites.tufts.edu/neudc2017/files/2017/10/paper_417.pdf

Bruhn, Miriam and David McKenzie. 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics* 1 (4): 200-232.

Cahyadi, Nur, Rema Hanna, Benjamin A. Olken, Rizal Adi Prima, Elan Satriawan, and Ekki Syamsulhakim. 2018. Cumulative Impacts of Conditional Cash Transfer Programs: Experimental Evidence from Indonesia. *NBER Working Paper 24670.* National Bureau of Economic Research.

Canning, David, Sangeeta Raja, and Abdo S. Yazbeck. 2015. Africa's Demographic Transition: Dividend or Disaster? Africa Development Forum; Washington, DC: World Bank; and Agence Française de Développement.

Chay, Kenneth Y and Michael Greenstone. 2003. The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. *Quarterly Journal of Economics* 118 (3): 1121-1167.

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz. 2016. The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review* 106 (4): 855-902.

Currie, Janet and Douglas Almond. 2011. Human capital development before age five. *Handbook of Labor Economics* 4:1315-1486.

Dahl, Gordon B and Lance Lochner. 2012. The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit. *American Economic Review* 102 (5): 1927-1956.

de Brauw, Alan and John Hoddinott. 2011. Must Conditional Cash Transfer Programs be Conditioned to be Effective? The Impact of Conditioning Transfers on School Enrollment in Mexico. *Journal of Development Economics* 96(2): 359-370.

Duflo, Esther. 2012. Women Empowerment and Economic Development. *Journal of Economic Literature* 50 (4): 1051-1079.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. Education, HIV, and Early Fertility: Experimental Evidence from Kenya. *American Economic Review* 105 (9): 2757-97.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2017. The Impact of Free Secondary Education: Experimental Evidence from Ghana. *Unpublished Manuscript*. https://web.stanford.edu/~pdupas/DDK_GhanaScholarships.pdf

Evans, David and Mũthoni Ngati. 2018. School Costs, Short-Run Participation, and Long-Run Outcomes: Evidence from Kenya. *World Bank Policy Research Working Paper*, no. 8421.

Fan, Jianqing. 1992. Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87 (420): 998-1004.

Filmer, Deon and Norbert Schady. 2014. The Medium-Term Effects of Scholarships in a Low-Income Country. *Journal of Human Resources* 49 (3): 663-694.

Fiszbein, Ariel, Norbert Schady, Francisco H.G. Ferreira, Margaret Grosh, Niall Keleher, Pedro Olinto, and Emmanuel Skoufias,. 2009. Conditional Cash Transfers: Reducing Present and Future Poverty. World Bank Policy Research Report. Washington, DC: World Bank.

Garcia, Sandra and Juan Saavedra. 2017. Educational Impacts and Cost-Effectiveness of Conditional Cash Transfer Programs in Developing Countries: A Meta-Analysis. *Review of Educational Research* 87(5): 921-965.

Gertler, Paul J, Sebastian W Martinez, and Marta Rubio-Codina. 2012. Investing Cash Transfers to Raise Long-Term Living Standards. *American Economic Journal: Applied Economics* 4 (1): 164-192.

Handa, Sudhanshu, Luisa Natali, David Seidenfeld, Gelson Tembo, and Benjamin Davis. 2018. Can unconditional cash transfers raise long-term living standards? Evidence from Zambia. *Journal of Development Economics* 133: 42-65.

Haushofer, Johannes and Jeremy Shapiro. 2016. The Short-Term Impact Of Unconditional Cash Transfers To The Poor: Experimental Evidence From Kenya. *Quarterly Journal of Economics* 131 (4): 1973-2042.

Haushofer, Johannes and Jeremy Shapiro. 2018. The Long-Term Impact Of Unconditional Cash Transfers: Experimental Evidence From Kenya. *Unpublished manuscript*. http://jeremypshapiro.com/papers/Haushofer_Shapiro_UCT2_2018-01-30_paper_only.pdf

Heath, Rachel and A. Mushfiq Mobarak. 2015. Manufacturing Growth and the Lives of Bangladeshi Women. *Journal of Development Economics* 115:1-15.

Heckman, James J. and Chase O. Corbin. 2016. Capabilities and Skills. *Journal of Human Development and Capabilities* 17 (3): 342-359.

Heckman, James J and Stefano Mosso. 2014. The Economics of Human Development and Social Mobility. *NBER Working Paper 19925*. National Bureau of Economic Research.

Heckman, James J and Tim Kautz. 2013. Fostering and Measuring Skills: Interventions that Improve Character and Cognition. *NBER Working Paper 19656*. National Bureau of Economic Research.

Hicks, Joan Hamory, Michael Kremer, Isaac Mbiti, and Edward Miguel. 2018. Addressing Youth Unemployment through Training and Grants: Experimental Evidence from Kenya. ASSA 2018 Presentation.

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond. 2016. Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review* 106 (4): 903-934.

IFPRI. 2014. Global nutrition report 2014: Actions and accountability to accelerate the world's progress on nutrition. Washington, DC: International Food Policy Research Institute (IFPRI).

Jensen, Robert. 2012. Do Labor Market Opportunities affect Young Women's Work and Family Decisions? Experimental Evidence from India. *Quarterly Journal of Economics* 127 (2): 753-792.

Kling, J., and J. Liebman, ‘‘Experimental Analysis of Neighborhood Effects on Youth,’’ Unpublished manuscript (2004).

LaFave, Daniel and Duncan Thomas. 2016. Height and Cognition at Work: Labor Market Productivity in a Low Income Setting. *NBER Working Paper 22290.* National Bureau of Economic Research.

Lee, David S. 2009. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies* 76 (3): 1071-1102.

Levine, Ruth, Cynthia Lloyd, Margaret Greene, and Caren Grown. 2008. Girls Count: A Global Investment and Action Agenda. Washington, D.C.: Center for Global Development.

Lloyd, Cynthia B. and Juliet Young. 2009. New Lessons: The Power of Educating Adolescent Girls: A Girls Count Report on Adolescent Girls. The Population Council, Inc. http://www.popcouncil.org/uploads/pdfs/2009PGY_NewLessons.pdf.

Manley, James, Seth Gitter, and Vanya Slavchevska. 2013. How Effective are Cash Transfers at Improving Nutritional Status? *World Development* 48:133-155.

Milligan, Kevin and Mark Stabile. 2009. Do Child Tax Benefits Affect the Wellbeing of Children? Evidence from Canadian Child Benefit Expansions. *American Economic Review* 99 (2): 128-132.

Molina-Millan, Teresa, Tania Barham, Karen Macours, John A. Maluccio, and Marco Stampini. 2018a. Long-Term Impacts of Conditional Cash Transfers: Review of the Evidence *World Bank Research Observer*, *forthcoming*.

Molina-Millan, Teresa, Karen Macours, John A. Maluccio, and Luis Tejerina. 2018b. Experimental Long-term Effects of Early Childhood and School-age Exposure to a Conditional Cash Transfer Program. *Working paper.*

Molyneux, Maxine, With Nicola Jones, and Fiona Samuels. 2016. Can Cash Transfer Programmes Have ‘Transformative’ Effects? *The Journal of Development Studies* 52 (8): 1087-1098.

Munshi, Kaivan D. and Mark R. Rosenzweig. 2006. Traditional Institutions Meet the Modern World: Caste, Gender and Schooling Choice in a Globalizing Economy. *American Economic Review* 96 (4): 1225-1252.

National Statistical Office and ICF Macro. 2005. Malawi Demographic and Health Survey 2010

Naudeau, Sophie, Rifat Hasan, and Anne Bakilana. 2015. Adolescent Girls in Zambia: Introduction and Overview. Policy Brief: Zambia:

http://elibrary.worldbank.org/doi/abs/10.1596/24597.

Oster, Emily and Bryce Millett Steinberg. 2013. Do IT Service Centers Promote School Enrollment? Evidence from India. *Journal of Development Economics* 104:123-135.

Ozier, Owen. 2015. The Impact of Secondary Schooling in Kenya: A Regression Discontinuity Analysis. *World Bank Policy Research Working Paper*, no. 7384.

Parker, Susan and Tom S. Vogl. 2018. Do Conditional Cash Transfers Improve Economic Outcomes in the Next Generation? *NBER Working Paper 24303*. National Bureau of Economic Research.

Schady, Norbert R. and Maria Caridad Araujo. 2008. Cash Transfers, Conditions, and School Enrollment in Ecuador. *Economia*, 8, 43–70.

Thomas, Duncan, John Strauss, and Maria-Helena Henriques. 1991. How Does Mother's Education Affect Child Height? *Journal of Human Resources* 26 (2): 183-211.

World Bank. 2010. World Development Indicators 2010. World Development Indicators. World Bank.

Zomba City Assembly. 2009. Zomba district socio economic profile 2009-2012.

Table I: Baseline means and balance

	Baseline Dropout		Baseline Schoolgirl			
	Mean (s.d.)		Mean (s.d.)			p-value (CCT-UCT)
	Control group	Conditional group	Control group	Conditional group	Unconditional Group	
	(1)	(2)	(3)	(4)	(5)	(6)
Urban Household	0.181	0.129	0.346	0.478	0.418	0.726
	(0.385)	(0.335)	(0.476)	(0.500)	(0.494)	
Mother Alive	0.783	0.749	0.839	0.800	0.828	0.431
	(0.413)	(0.434)	(0.368)	(0.401)	(0.378)	
Father Alive	0.656	0.649	0.709	0.718	0.76	0.341
	(0.476)	(0.478)	(0.454)	(0.451)	(0.428)	
Household Size	6.120	6.104	6.375	6.341	6.659	0.156
	(2.388)	(2.617)	(2.262)	(2.134)	(2.063)	
Asset Index	-0.831	-0.743	0.632	1.100	1.373*	0.572
	(2.233)	(2.484)	(2.575)	(2.721)	(2.444)	
Age	17.579	17.162	15.228	14.919	15.466	0.002
	(2.397)	(2.478)	(1.904)	(1.828)	(1.926)	
Highest Grade Attended	6.105	5.940	7.506	7.262	7.928**	0.004
	(2.856)	(2.864)	(1.651)	(1.601)	(1.587)	
Never Had Sex	0.315	0.294	0.800	0.807	0.790	0.682
	(0.465)	(0.456)	(0.400)	(0.395)	(0.408)	
Ever Pregnant	0.445	0.420	0.021	0.029	0.029	0.964
	(0.498)	(0.494)	(0.144)	(0.169)	(0.168)	
Chi-squared joint test of orthogonality (p-value)		0.168		0.122	0.121	0.032

Notes: Mean differences statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence. Stars on the coefficients in columns (2) indicate significantly different than the control group for baseline dropouts.Stars on the coefficients in column (4) and (5) indicate significantly different than the control group for baseline schoolgirls. Means are weighted to make them representative of the target population in the study EAs.

Table II: Attrition

	Baseline Dropout		Baseline Schoolgirl	
	=1 if Completed Household Survey Round 4			
	(1)	(2)	(3)	(4)
=1 if Conditional	-0.007	-0.008	0.055***	0.056***
	(0.031)	(0.029)	(0.019)	(0.018)
=1 if Unconditional			0.058***	0.061***
			(0.023)	(0.021)
p-value UCT vs. CCT	-	-	0.896	0.825
p-value Treatment	0.828	0.774	0.004	0.002
Baseline controls interacted with treatment?	NO	YES	NO	YES
p-value on joint F-test for interactions CCT	-	0.009	-	0.332
p-value on joint F-test for interactions UCT	-	-	-	0.101
p-value UCT interactions vs. CCT interactions	-	-	-	0.690
Mean in Control Group	0.843	0.843	0.875	0.875
Number of observations	885	885	2,273	2,273

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. All regressions include baseline centered values of the following variables: age indicators, stratum indicators, household asset index, highest grade attended, an indicator for never had sex. Columns (2) and (4) interact the centered baseline controls with treatment. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table III: Program impacts on education and learning (beneficiaries)

Panel A: Baseline Dropouts	Highest Grade Completed			English Test Score (Standardized)	TIMMS Math Score (Standardized)	Non-TIMMS Math Score (Standardized)	Cognitive Test Score (Standardized)	Competencies Score (Standardized)
	During Program	End of Program	Two Years After Program	End of Program				Two Years After Program
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
=1 if Conditional Schoolgirl	0.579***	0.558***	0.621***	0.079	0.147***	0.116	0.163**	0.064
	(0.073)	(0.102)	(0.125)	(0.071)	(0.056)	(0.072)	(0.070)	(0.057)
Mean in Control Group	6.345	6.967	6.997	0.000	0.000	0.000	0.000	0.000
Sample Size	697	718	744	704	704	704	704	742
Panel B: Baseline Schoolgirls								
=1 if Conditional Schoolgirl	0.078	0.126*	0.120	0.148***	0.136**	0.068	0.181***	0.065
	(0.090)	(0.069)	(0.080)	(0.056)	(0.069)	(0.063)	(0.050)	(0.058)
=1 if Unconditional Schoolgirl	0.122	0.103	0.095	-0.068	-0.027	0.026	0.094	0.098
	(0.109)	(0.121)	(0.129)	(0.090)	(0.106)	(0.090)	(0.129)	(0.067)
p-value UCT vs. CCT	0.708	0.854	0.850	0.035	0.157	0.657	0.514	0.630
p-value Treatment	0.469	0.174	0.309	0.021	0.118	0.560	0.002	0.297
Mean in Control Group	8.590	9.677	10.415	0.000	0.000	0.000	0.000	0.000
Sample Size	1,965	2,019	2,049	2,000	2,000	2,000	2,000	2,048

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. The cognitive test score is based on Raven's Colored Progressive Matrices. Math and English reading comprehension tests were developed based on the Malawian school curricula. Five questions (four from the Fourth Grade test and one from the Eighth Grade test) from Trends in Mathematics and Science Study (TIMMS) 2007, which is a cycle of internationally comparative assessments in mathematics and science carried out at the fourth and eighth grades every four years, were added to the math test. Competencies represent a set of skills that were anticipated to be sensitive to education and relevant for non-formal employment. The skills tested included reading and following instructions to apply fertilizer; making correct change during hypothetical market transactions; sending text messages and using the calculator on a mobile phone, and calculating profits under hypothetical business scenarios. All test scores and the competency index were standardized to have a mean of zero and a standard deviation of one in the control group. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, an indicator for never had sex, and whether the respondent participa+A1:I29ted in the pilot phase of the development of the testing instruments. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Note that in Rounds 2 and 3, highest grade *completed* is actually highest grade *attended*. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table IV: Program impacts on marriage and fertility (beneficiaries)

Panel A: Baseline Dropouts	=1 if Ever Married			Age First Marriage	=1 if Ever Pregnant			Number of Live Births			Age at First Birth	Desired Fertility
	During Program	End of Program	Two Years After Program	Two Years After Program	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	Two Years After Program	Two Years After Program
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
=1 if Conditional Schoolgirl	-0.140***	-0.157***	-0.107***	0.431***	-0.057*	-0.081***	-0.040*	-0.005	-0.095**	-0.147***	0.272*	-0.172*
	(0.029)	(0.037)	(0.032)	(0.155)	(0.030)	(0.027)	(0.021)	(0.033)	(0.044)	(0.054)	(0.164)	(0.087)
Mean in Control Group	0.291	0.575	0.809	19.644	0.610	0.784	0.924	0.520	0.819	1.380	18.499	3.217
Sample Size	698	718	744	500	698	718	744	698	718	744	634	744
Panel B: Baseline Schoolgirls												
=1 if Conditional Schoolgirl	0.000	-0.010	-0.035	-0.011	0.008	0.027	-0.024	0.023*	0.003	0.020	-0.144	-0.072
	(0.012)	(0.024)	(0.027)	(0.148)	(0.015)	(0.027)	(0.034)	(0.014)	(0.022)	(0.036)	(0.136)	(0.064)
=1 if Unconditional Schoolgirl	-0.033***	-0.083***	-0.010	0.486**	-0.013	-0.063**	-0.001	0.013	-0.055*	-0.024	0.001	-0.017
	(0.012)	(0.024)	(0.046)	(0.200)	(0.017)	(0.028)	(0.042)	(0.017)	(0.030)	(0.046)	(0.168)	(0.056)
p-value UCT vs. CCT	0.026	0.018	0.613	0.032	0.314	0.009	0.614	0.641	0.075	0.410	0.436	0.477
p-value Treatment	0.023	0.004	0.448	0.050	0.600	0.025	0.760	0.209	0.151	0.705	0.547	0.533
Mean in Control Group	0.047	0.180	0.402	18.651	0.092	0.247	0.501	0.055	0.199	0.511	18.718	2.974
Sample Size	1,967	2,018	2,049	821	1,966	2,019	2,049	1,966	2,019	2,049	998	2,048

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. We correct for inconsistencies in 'ever married' and 'ever pregnant' across rounds.All regressions are weighted to make them representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table V: Program impacts on HIV and Anemia (beneficiaries)

	=1 if HIV Positive			=1 if Anemic
	During Program	End of Program	Two Years After Program	Two Years After Program
	(1)	(2)	(3)	(4)
Panel A: Baseline Dropouts				
=1 if Conditional Schoolgirl	0.022	0.020	0.012	0.039
	(0.024)	(0.023)	(0.026)	(0.035)
Mean in Control Group	0.06	0.094	0.135	0.255
Sample Size	373	694	715	711
Panel B: Baseline Schoolgirls				
=1 if Conditional Schoolgirl	-0.020**	-0.003	-0.001	0.012
	(0.009)	(0.011)	(0.019)	(0.031)
=1 if Unconditional Schoolgirl	-0.015	-0.019*	-0.002	-0.065*
	(0.012)	(0.012)	(0.023)	(0.033)
p-value UCT vs. CCT	0.616	0.237	0.980	0.068
p-value Treatment	0.112	0.249	0.996	0.122
Mean in Control Group	0.026	0.035	0.055	0.243
Sample Size	1,192	2,002	1,977	1,979

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. An individual is considered anemic if her hemoglobin count is less than or equal to 11g/dL if pregnant and less than or equal to 12d/dL if non-pregnant based on WHO guidelines to define mild anemia. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table VI: Program impacts on labor market outcomes and empowerment (beneficiaries: **primary** outcomes)

Panel A: Baseline Dropouts	Labor Market Outcomes			Empowerment				
	Opportunity Cost of Time (2012 USD)	Typical Wage in Past Three Months (2012 USD)	Proportion of Hours Spent in Self-Employment or Paid Work in Past Week	Super-Index of Overall Empowerment (Standardized)	Change in Subjective Wellbeing from Five Years Ago to Today	Super-Index of Unmarried Empowerment (Standardized)	Super-Index of Married Empowerment (Standardized)	Married Index of Economic Control (Standardized)
	Two Years After Program							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
=1 if Conditional Schoolgirl	-0.037	-0.140**	-0.011	-0.083	-0.032	0.018	-0.113	-0.118
	(0.079)	(0.068)	(0.009)	(0.074)	(0.232)	(0.112)	(0.102)	(0.096)
Mean in Control Group	0.707	0.375	0.061	0.000	1.120	0.000	0.000	0.000
Sample Size	718	743	744	744	744	289	455	455
Panel B: Baseline Schoolgirls								
=1 if Conditional Schoolgirl	-0.051	-0.011	0.003	0.049	0.276	0.111	0.068	-0.107
	(0.101)	(0.058)	(0.005)	(0.082)	(0.187)	(0.098)	(0.095)	(0.108)
=1 if Unconditional Schoolgirl	-0.115	0.036	0.002	-0.159*	0.176	-0.094	-0.342***	0.147
	(0.074)	(0.104)	(0.008)	(0.081)	(0.190)	(0.109)	(0.099)	(0.307)
p-value UCT vs. CCT	0.550	0.665	0.842	0.052	0.650	0.120	0.001	0.406
p-value Treatment	0.297	0.910	0.784	0.101	0.306	0.287	0.001	0.484
Mean in Control Group	0.897	0.212	0.029	0.000	0.906	0.000	0.000	0.000
Sample Size	2,002	2,048	2,045	2,049	2,049	1,271	776	774

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Opportunity cost of time is calculated by taking the minimum daily wage the respondent would take for one year of work in her village. Detail on the construction of the super-indices can be found at https://drive.google.com/file/d/1hvI79ltywocFr-pafqz8_Dtg2ZXNhcHd/view. The change in subjective wellbeing asks the respondent where she sees herself on a 10-step ladder comparing five years ago to today, where zero represents the worst possible life she could have and 10 represents the best possible life she could have. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table VII: Primary outcomes with multiple testing adjustments (original p-values and FDR q-values)

	Baseline Schoolgirl						Baseline Dropout	
Outcomes	CCT vs. Control		UCT vs. Control		CCT vs. UCT		CCT vs. Control	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	p-value	q-value	p-value	q-value	p-value	q-value	p-value	q-value
Highest Grade Completed	0.136	1.000	0.465	1.000	0.850	1.000	0.000	0.007
Competencies Score (Standardized)	0.269	1.000	0.147	0.478	0.630	1.000	0.263	0.237
=1 if Ever Married	0.206	1.000	0.829	1.000	0.613	1.000	0.001	0.007
Age at First Marriage	0.940	1.000	0.016	0.289	0.032	0.465	0.006	0.022
=1 if Ever Pregnant	0.471	1.000	0.980	1.000	0.614	1.000	0.054	0.099
Number of Live Births	0.580	1.000	0.600	1.000	0.410	1.000	0.007	0.022
Age at First Birth	0.292	1.000	0.997	1.000	0.436	1.000	0.100	0.145
= if HIV Positive	0.955	1.000	0.938	1.000	0.980	1.000	0.649	0.504
=1 if Anemic	0.699	1.000	0.053	0.299	0.068	0.465	0.263	0.237
Opportunity Cost of Time	0.617	1.000	0.120	0.478	0.550	1.000	0.641	0.504
Typical Daily Wage in Last Three Months	0.847	1.000	0.726	1.000	0.665	1.000	0.041	0.090
Proportion of Hours Spent in Self-Employment or Paid Work in Past	0.502	1.000	0.831	1.000	0.842	1.000	0.221	0.237
Super Index of Overall Empowerment (Standardized)	0.551	1.000	0.051	0.299	0.052	0.465	0.890	0.504
Change in Subjective Wellbeing from Five Years Ago to Today	0.143	1.000	0.354	1.000	0.650	1.000	0.263	0.237

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, an indicator for never had sex, and whether the respondent participated in the pilot phase of the development of the testing instruments. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table VIII: Program impacts on height-for-age z-scores (children of beneficiaries**: baseline dropouts**)

	Raw Effect				**Direct Effect**
Panel A: Born During Program	Gender	+ Maternal Selection weights	+ Paternal Selection Controls	+ Child Age	+ Mother Age
	(1)	(2)	(3)	(4)	(5)
=1 if Conditional Schoolgirl	-0.015	-0.174	-0.139	-0.154	-0.051
	(0.128)	(0.149)	(0.143)	(0.140)	(0.136)
Sample Size	367	367	367	367	367
Panel B: Born Within 9 Months of Program Ended					
=1 if Conditional Schoolgirl	0.353	0.518*	0.394	0.411*	0.577**
	(0.296)	(0.303)	(0.249)	(0.234)	(0.260)
Sample Size	88	88	88	88	88
Panel C: Born More than 9 Months After Program Ended					
=1 if Conditional Schoolgirl	-0.269	-0.175	-0.127	-0.137	-0.183
	(0.168)	(0.192)	(0.161)	(0.154)	(0.152)
Sample Size	287	287	287	287	287
Control Structure:					
Maternal selection controls + propensity weight		X	X	X	X
Father selection controls			X	X	X
Cubic in child age in months				X	X
Maternal age in years, age interactions					X

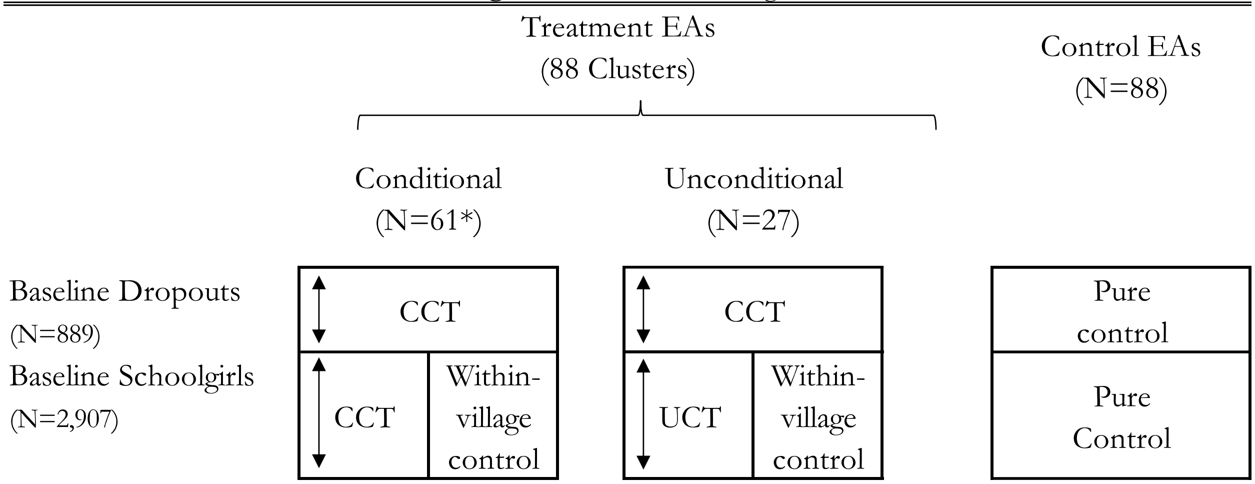
Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. The height-for-age z-score is calculated using the 2006 WHO child growth standards. Specification (1) controls for the gender of the child. Specification (2) adds selection weights and controls directly for maternal baseline characteristics (stratum indicators, household asset index, highest grade attended, and an indicator for never had sex). Specification (3) adds controls for paternal attributes (highest education level, religion, ethnicity, main activity, and likely HIV status). Specification (4) adds a linear, quadratic, and cubic in child age. Specification (5) adds maternal age and maternal age interacted with the other baseline covariates. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table IX: Program impacts on height-for-age z-scores (children of beneficiaries: **baseline schoolgirls**)

	Raw Effect				**Direct Effect**
Panel A: Born During Program	Gender	+ Maternal Selection weights	+ Paternal Selection Controls	+ Child Age	+ Mother Age
	(1)	(2)	(3)	(4)	(5)
=1 if Conditional Schoolgirl	0.155	-0.050	-0.054	0.023	0.124
	(0.162)	(0.192)	(0.186)	(0.177)	(0.155)
=1 if Unconditional Schoolgirl	0.953**	0.525**	0.549*	0.666**	0.523*
	(0.476)	(0.221)	(0.306)	(0.315)	(0.299)
p-value UCT vs. CCT	0.091	0.022	0.028	0.024	0.115
p-value Treatment	0.123	0.040	0.089	0.072	0.218
Sample Size	315	315	315	315	315
Panel B: Born Within 9 Months of Program Ended					
=1 if Conditional Schoolgirl	0.251	0.156	0.235	0.125	0.086
	(0.279)	(0.263)	(0.240)	(0.175)	(0.194)
=1 if Unconditional Schoolgirl	0.177	0.163	0.109	-0.431**	-0.434**
	(0.514)	(0.315)	(0.336)	(0.183)	(0.193)
p-value UCT vs. CCT	0.887	0.984	0.725	0.013	0.028
p-value Treatment	0.663	0.787	0.619	0.028	0.047
Sample Size	214	211	211	211	211
Panel C: Born More than 9 Months After Program Ended					
=1 if Conditional Schoolgirl	-0.011	0.497	0.149	0.264	0.257
	(0.187)	(0.445)	(0.199)	(0.196)	(0.179)
=1 if Unconditional Schoolgirl	-0.351**	-0.651***	-0.336	-0.102	-0.123
	(0.174)	(0.242)	(0.212)	(0.168)	(0.183)
p-value UCT vs. CCT	0.115	0.006	0.025	0.068	0.078
p-value Treatment	0.114	0.002	0.075	0.184	0.186
Sample Size	507	506	506	506	506
Control Structure:					
Maternal selection controls + propensity weight		X	X	X	X
Father selection controls			X	X	X
Cubic in child age in months				X	X
Maternal age in years, age interactions					X

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. The height-for-age z-score is calculated using the 2006 WHO child growth standards. Specification (1) controls for the gender of the child. Specification (2) adds selection weights and controls directly for maternal baseline characteristics (stratum indicators, household asset index, highest grade attended, and an indicator for never had sex). Specification (3) adds controls for paternal attributes (highest education level, religion, ethnicity, main activity, and likely HIV status). Specification (4) adds a linear, quadratic, and cubic in child age. Specification (5) adds maternal age and maternal age interacted with the other baseline covariates. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

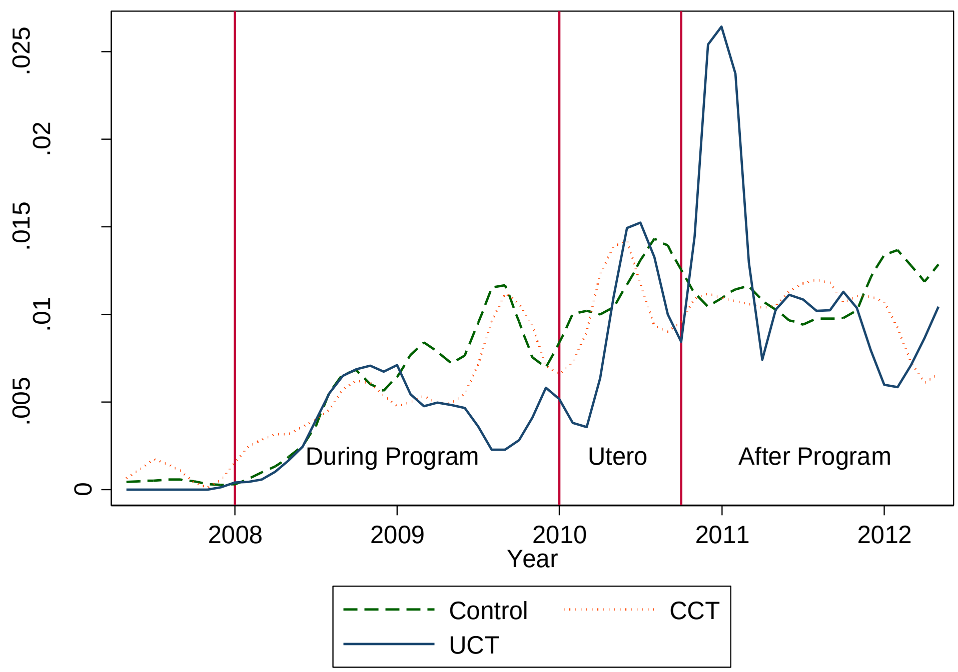
Figure I: Research Design



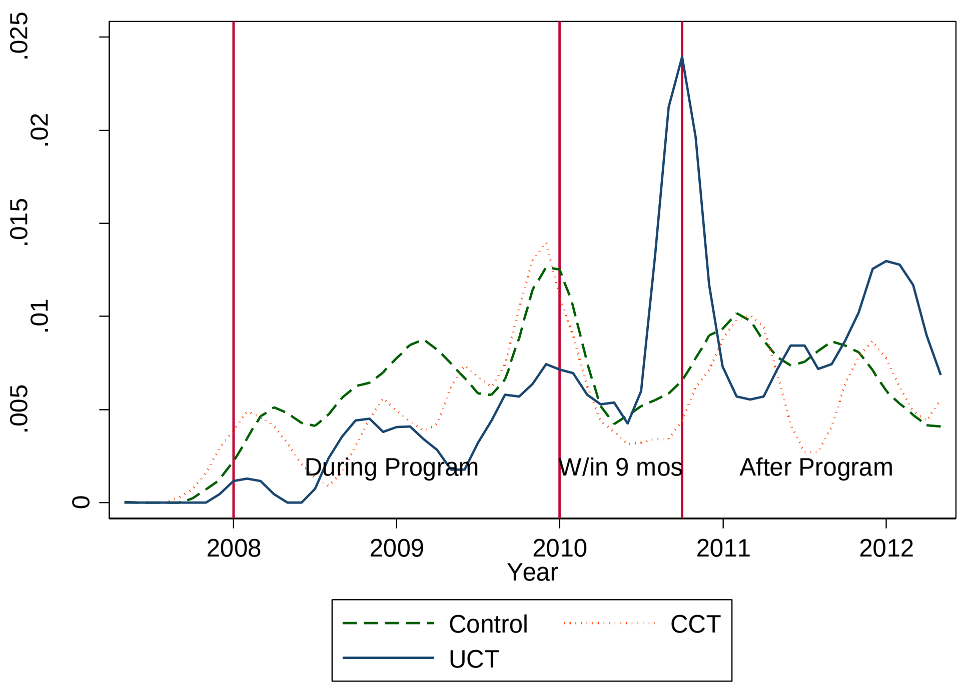
*In 15 of the 61 conditional treatment clusters only baseline dropouts were treated.

Figure II: Monthly marriage and fertility rates for baseline schoolgirls

Panel A: Monthly Fertility Rates among the Baseline Schoolgirls.



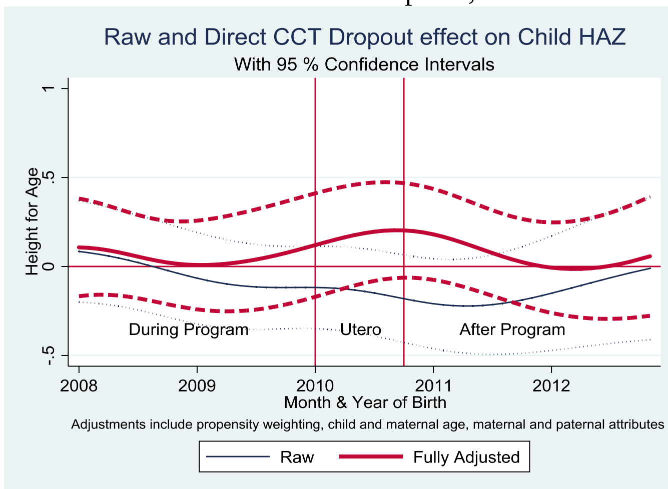
Panel B: Monthly Marriage Rates among the Baseline Schoolgirls



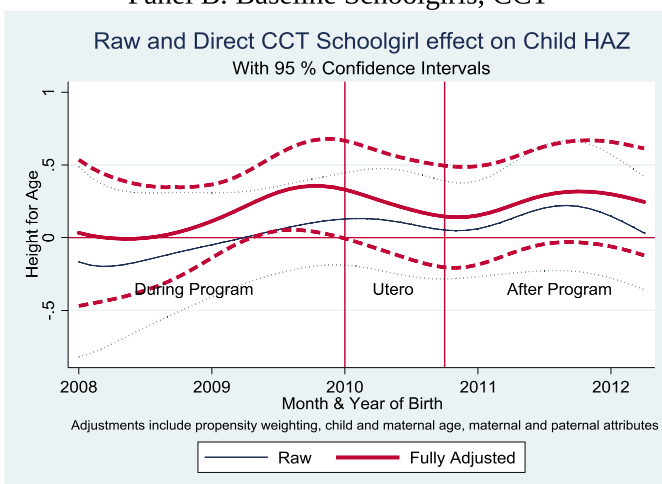
Notes: Figures illustrate the smoothed fraction of core respondents who give birth (Panel A) or get married (Panel B) in each month using retrospective information on the month of birth and marriage, respectively.

Figure III: Fan regressions of height-for-age z-scores by month of birth, raw and fully adjusted treatment effects with 95% confidence intervals

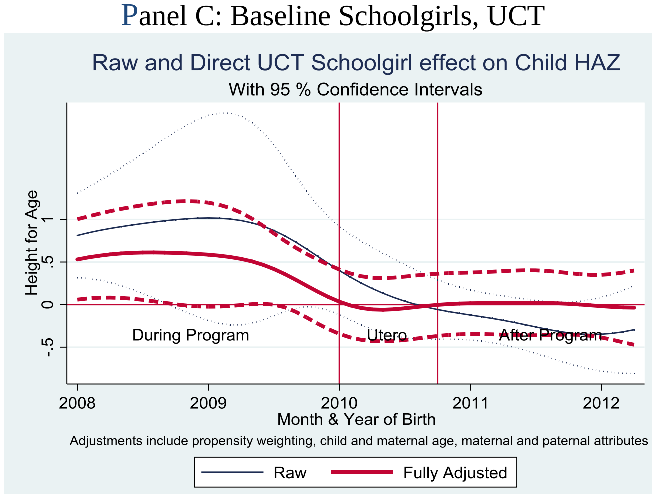
Panel A: Baseline Dropouts, CCT



Panel B: Baseline Schoolgirls, CCT



Panel C: Baseline Schoolgirls, UCT



FOR ONLINE PUBLICATION

Supplemental Online Appendix for:

When the Money Runs Out:

Do Cash Transfers Have Sustained Effects on Human Capital Accumulation?

Sarah Baird*, Craig McIntosh and Berk Özler

*Corresponding author. E-mail: sbaird@gwu.edu

March 30, 2019

This PDF file includes:

Appendix A: Estimation of Treatment Effects on Children
Tables S1 to S11

Appendix A: Estimation of Treatment Effects on Children

This appendix provides an overview of the empirical issues involved in estimating treatment effects on child outcomes when the intervention under investigation targets prospective mothers and starts prior to their pregnancies. As the intervention may have altered the composition of children subsequently observed, we suggest a simple sequence of assumptions and steps in an attempt to move from the total reduced-form effect of the intervention to a more standard causal effect on the children actually born.

The natural experimental literature has recognized that maternal selection and differential mortality represent plausible causal pathways when analyzing child outcomes. Using data from the US, Buckles and Hungerman (2013) show that nearly half of the well-documented effects of season of birth on later-life outcomes can be explained by variation in the types of women giving birth across seasons. Aaronson, Lange and Mazumder (2014) show extensive margin selection also contributes to the quantity-quality tradeoff for children. This phenomenon is well documented in the developed world, and yet Currie and Vogl (2012) suggest that these extensive margin effects are likely to be more pronounced in the developing world where differential mortality, as well as differential fertility, is an operative channel. It has now become standard in the natural experiments literature to test for the presence of selection effects when studying child outcomes (see, for example, Almond 2006; Adhvaryu et al. 2016; Black et al. 2017). In many cases in this literature, the obvious selection effect suggests that the simple casual effects are lower bounds, but we present a context in which selection and treatment effects go in the same direction. It is therefore critical to attempt to control for the selection mechanism as a mediator, to isolate how much of the observed effect may have a simple causal interpretation.

To use counterfactual outcomes to represent impacts on subsequent children, we must define the universe as being the *potential children*: all those who might exist under either the treated or control state.[40] Outcome measured for potential child *i* at time *t* is Y_{it}. This outcome is only observed if the child is born and survives, a binary outcome denoted by $S_{it} = 1$. The probability that woman *i* has a surviving child at time *t*, as a function of her baseline characteristics, can be

[40] This study tracks female respondents and their descendants, so it is concerned with the potential children of a fixed set of women. We do not attempt to capture all potential children born to the fathers in this study, and to do so would have needed to specify a group of males at baseline and tracked them. The control for father characteristics in this context therefore sits behind a layer of female selection, and so the assumptions needed to control for father type are stronger than those needed to control for mother type.

written as $S_{it} = S(X_i)$, where X_i is a set of pre-treatment maternal characteristics. For children who are born, we observe

$$Y_{it} = Y_{it}(x_{it}, a_{it}; X_i, A_{it}, Z_{it} \mid S_{it} = 1)$$

where A_{it} is the age of the mother at the time at which child data is collected, x_{it} are child-level determinants of the outcome, a_{it} is child age, and Z_{it} gives attributes of the father.

Critically, in a study tracking potential mothers, X_i is observed for all respondents (regardless of whether they had children) and hence can be used to predict fertility within the universe of these potential mothers. Controlling for selection into motherhood thus resembles standard attrition adjustment and requires a (weaker) selection on observables assumption. Controlling for factors that are observed only among extant children, however, must be done on the intensive margin, and so resembles mediation analysis as in Baron and Kenny (1986), which requires substantially stronger assumptions. Specifically, we must now assume that there is a globally correct functional form across both treatment and control, so that inclusion of the mechanism controls does not open a 'backdoor path' between the mediator and some other, unobserved determinant of outcomes (for more discussion of these assumptions, see Sobel 2008; Flores and Flores-Lagunes 2009; Bullock, Green and Ha 2010; Imai, Tingley and Yamamoto 2013; Huber 2014; and Heckman and Pinto 2015).

Given these strong assumptions, we now walk through a set of distinct treatment effects that we may wish to estimate, each of which has a different causal interpretation. We start with the simple reduced-form impact of the treatment on child outcomes and proceed to add successively stronger controls until we have isolated the *ceteris* paribus treatment effect that would have been observed had an experiment been conducted on the sample of children actually observed, rather than their mothers. Appendix Figure A1 presents a conceptual framework.

Total effect:

$$(1) \qquad E\left(Y_{it}^{1} - Y_{it}^{0} \mid S_{it} = 1\right).$$

This is the simple difference in outcomes between the children of those exposed to the treatment versus the control.

Correcting for Maternal Type Selection:

We can begin to control for the extensive margin effects of the treatment by modeling the probability that a child is born to a mother *i* during epoch *t* as: $\Pr(S_{it}=1)=\Phi(X_i,T_i)+\varepsilon_{it}$. This problem is exactly analogous to attrition, and so we can exploit the familiar toolkit to test and correct for it (Hanson 1978, Rosenbaum and Rubin 1983). Observational correction can be conducted using a probit model that regresses a binary indicator for giving birth during an epoch on a rich set of baseline covariates, a treatment indicator, and the interaction between treatment and the covariates. We can use this regression to predict the probability of birth for all core respondents by epoch, and weight the subsequent analysis by the inverse of this probability. This is the application of standard attrition-based inverse probability weighting to the fertility problem. The required assumption is that there be no unobserved determinants of fertility that are correlated with the treatment or the treatment*covariate interactions. Regressions weighted by $1/\hat{\Pr}(S_{it}=1)$, subject to this assumption, are now representative of the entire original sample of core respondents and hence not subject to selection effects arising from the decision to give birth or not. We can also use OLS to control for the same set of maternal baseline characteristics X_i to provide estimates that are "doubly robust" to the extensive margin selection controls (Robins and Rotnitzky 2001; Van der Laan and Robins 2003; Bang and Robins 2005). This then provides an estimate of the impact on children if the composition of women who gave birth was identical in treatment and control in each epoch:

$$(2)\qquad E\left(Y_{it}^{1}-Y_{it}^{0}\mid X=\bar{X},\Pr(S=1\mid T)=\bar{S}\right)$$

Correcting for Paternal Type Selection:

The next selection margin is father type. In our data structure we do not observe the attributes of the universe of potential fathers, and rather have data on the fathers of children actually born. We therefore must control for paternal characteristics on the intensive margin rather than using the selection IPW approach that we use for mothers. The assumptions underlying these paternal controls are therefore the strong assumptions of mediation analysis, rather than those of attrition propensity weighting. Subject to these assumptions, the inclusion of paternal covariates gives us the expected treatment-control difference holding both maternal and paternal types constant across the treatment and control groups.

$$(3) \qquad E\left(Y_{it}^{1} - Y_{it}^{0} \mid X = \bar{X}, Z(T) = \bar{Z}, \Pr(S=1 \mid T) = \bar{S}\right)$$

Correcting for Child Age:

Differences in the composition of child age across treatment and control can lead to large differences that are, in fact, completely trivial. If, for example, the treatment led to a delay in births, then the treatment children will be younger on average and hence may perform more poorly on a broad range of tasks than the control children.[41] We can recover a meaningful treatment effect by comparing children in treatment and control *at the same age.* We achieve this by including linear, quadratic, and cubic controls for child age in months in our regressions:

$$(4) \qquad E\left(Y_{it}^{1} - Y_{it}^{0} \mid a(T) = \bar{a}, X = \bar{X}, Z(T) = \bar{Z}, \Pr(S=1 \mid T) = \bar{S}\right)$$

Direct Treatment Effect:

Maternal age represents a potentially important mechanism for improvements in child outcomes for the same set of mothers, even though it represents an extensive margin effect in that changes in maternal age must necessarily lead to a different set of children being born. It is possible, for example, that an intervention that, all else equal, simply delayed fertility from age 13 to age 18 would lead to improved child outcomes due to increased gynecological maturity. Effects driven only by changes in age therefore have a meaningful causal interpretation that can be seen as *ceteris paribus* for mothers even though it operates on the extensive margin for children. To control for age as a mediating variable, we can include age and $A*X$ interactions as covariates.

Subject to strong assumptions of (i) selection on observables in the fertility equation, and (ii) correct functional form and common support in the Barron-Kenny controls for the mechanisms, the resulting adjusted difference provides the 'direct' treatment effect of the program on a sample of children made homogeneous across treatment and control by reweighting and regression adjustments. The result is a suggestive answer to an obvious policy question: "Does the intervention confer a protective effect on a given child?"

$$(5) \qquad E\left(Y_{it}^{1} - Y_{it}^{0} \mid a(T) = \bar{a}, Z(T) = \bar{Z}, X = \bar{X}, A(T) = \bar{A}, \Pr(S=1 \mid T) = \bar{S}\right)$$

[41] Conversely, as shown in Appendix Figure A2, the mean height-for-age z-score in our control group starts out very close to the mean of the reference group at birth, but declines steadily and rapidly as children get older, ending up almost two standard deviations below the global distribution by the time they are 36 months old. This seems to be a common pattern in poor countries (see, for example, Figure 1 in Barham, Macours and Maluccio 2013). Hence, comparing a younger cohort of children in the treatment group to an older cohort in control would spuriously show lower stunting in the treatment group in the absence of any meaningful effects on height.

Figure A1: Conceptual Framework for Causal Pathways to Child Treatment Effects

T_i : Treatment

Mother:

$S_{it}(X,T)=1$ Birth Outcome: Is a child born in this interval?

Maternal Selection effects that alter the identity of the woman who gives birth

$A_{it}^{C}(T)$: Birth Timing: Child Age

$A_{it}^{M}(T)$: Birth Timing: Maternal Age

Father

$Z_i(T)$: Father Selection

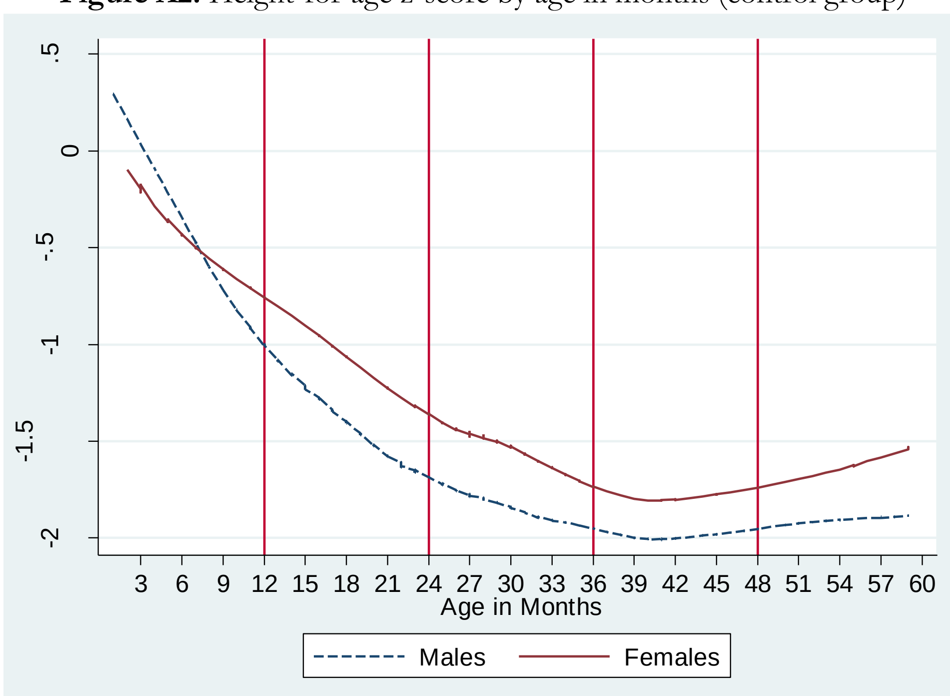
Child Selection effects that alter the identity of the child who is born

Child

D1. Direct impacts

Direct Treatment effect on the child actually born

Figure A2. Height-for-age z-score by age in months (control group)



References (for Appendix A)

Aaronson, Daniel, Fabian Lange, and Bhashkar Mazumder. 2014. Fertility Transitions Along the Extensive and Intensive margins. *American Economic Review* 104 (11): 3701-3724.

Adhvaryu, Achyuta, Prashant Bharadwaj, James Fenske, Anant Nyshadham, and Richard Stanley. 2016. Dust and Death: Evidence from the West African Harmattan. *CSAE Working Paper WPS/2016-03.* Centre for the Study of African Economies, University of Oxford.

Almond, D. 2006. Is the 1918 influenza pandemic over? Long-term effects of *In Utero* Influenza Exposure in the Post-1940 U.S. Population. *Journal of Political Economy* 114 (4): 672-712.

Baird, Sarah, Craig McIntosh, and Berk Özler. 2011. Cash or Condition? Evidence from a Cash Transfer Experiment. *Quarterly Journal of Economics* 126 (4): 1709-1753.

Baird, Sarah J, Richard S Garfein, Craig T McIntosh, and Berk Özler. 2012. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. *The Lancet* 379 (9823): 1320-1329.

Baird, Sarah, Jacobus De Hoop, and Berk Özler. 2013. Income shocks and adolescent mental health. *Journal of Human Resources* 48 (2): 370-403.

Bang, Heejung and James M Robins. 2005. Doubly robust estimation in missing data and causal inference models. *Biometrics* 61 (4): 962-973.

Barham, Tania, Karen Macours, and John A Maluccio. 2013. Boys' Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages for Early Childhood Interventions. *American Economic Review* 103 (3): 467-471.

Baron, Reuben M and David A Kenny. 1986. The moderator--mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51 (6): 1173.

Black, Sandra E, Aline Bütikofer, P Devereux, and K Salvanes. 2017. This Is Only a Test? Long-Run and Intergenerational Impacts of Prenatal Exposure to Radioactive Fallout. *NBER Working Paper 18987.* National Bureau of Economic Research.

Buckles, Kasey S and Daniel M Hungerman. 2013. Season of birth and later outcomes: Old questions, new answers. *Review of Economics and Statistics* 95 (3): 711-724.

Bullock, John G, Donald P Green, and Shang E Ha. 2010. Yes, but whats the mechanism?(dont expect an easy answer). *Journal of Personality and Social Psychology* 98 (4): 550.

Currie, Janet and Tom Vogl. 2012. Early-life health and adult circumstance in developing countries. *NBER Working Paper 18371.* National Bureau of Economic Research. National Bureau of Economic Research.

Flores, Carlos A. and Alfonso Flores-Lagunes. 2009. Identification and Estimation of Causal Mechanisms and Net Effects of a Treatment under Unconfoundedness. *IZA Discussion Paper No. 4237.*

Hanson, Robert Harold. 1978. The current population survey: design and methodology

Heckman, James J and Rodrigo Pinto. 2015. Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric Reviews* 34 (1-2): 6-31.

Huber, Martin. 2014. Identifying causal mechanisms (primarily) based on inverse probability weighting. *Journal of Applied Econometrics* 29 (6): 920-943.

Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. 2013. Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176 (1): 5-51.

Robins, James M and Andrea Rotnitzky. 2001. COMMENTS. *Statistica Sinica* 920-936.

Rosenbaum, Paul R and Donald B Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70 (1): 41-55.

Sobel, Michael E. 2008. Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics* 33 (2): 230-251.

Van der Laan, Mark J and James M Robins. 2003. Unified methods for censored longitudinal data and causality. Springer Science & Business Media.

Table S1: Survey Tracking Rates

	Overall	Baseline Dropout		Baseline Schoolgirl		
		Mean (s.d.)		Mean (s.d.)		
		Control group	Conditional group	Control group	Conditional group	Unconditional Group
	(1)	(2)	(3)	(4)	(5)	(6)
Found Indicator	0.941	0.923	0.917	0.936	0.955	0.956
	(0.236)	(0.267)	(0.276)	(0.245)	(0.197)	(0.197)
Deceased Indicator	0.007	0.022	0.018	0.005	0.004	0.004
	(0.086)	(0.147)	(0.134)	(0.069)	(0.066)	(0.066)
Surveyed Indicator	0.891	0.843	0.837	0.875	0.933	0.930
	(0.312)	(0.364)	(0.370)	(0.311)	(0.250)	(0.256)

Notes: Found indicates that the respondent was located by the enumerator and either surveyed, refused or deceased.

Table S2: Probit Regressions that are the basis for Inverse Propensity Weighting

	Outcome variable: Successfully Surveyed in Round 4					
	Baseline Dropouts			Baseline Schoolgirls		
	Probit Marginal Effects	Standard Error	p-value	Probit Marginal Effects	Standard Error	p-value
	(1)	(2)	(3)	(4)	(5)	(6)
Conditional Treatment	-0.003	0.029	0.920	0.048	0.015	0.007
Unconditional Treatment				0.090	0.012	0.000
Baseline Age 14	-0.098	0.141	0.485	0.014	0.044	0.754
Baseline Age 15	-0.162	0.127	0.198	0.017	0.037	0.653
Baseline Age 16	-0.131	0.119	0.267	-0.015	0.040	0.708
Baseline Age 17	-0.095	0.133	0.472	0.037	0.039	0.342
Baseline Age 18	-0.077	0.137	0.569	0.019	0.042	0.647
Baseline Age 19	0.004	0.139	0.977	-0.050	0.046	0.280
Baseline Age 20+	0.018	0.148	0.901	-0.000	0.055	0.999
Semi-rural stratum	0.022	0.051	0.672	0.052	0.017	0.001
Rural stratum	-0.068	0.059	0.255	0.024	0.024	0.307
Highest grade at baseline	-0.018	0.008	0.023	-0.002	0.005	0.717
Asset index baseline	-0.015	0.008	0.051	0.002	0.005	0.723
Never had sex baseline	0.050	0.046	0.293	0.016	0.015	0.275
CCT*Age 14	0.127	0.164	0.436	-0.016	0.073	0.823
CCT*Age 15	0.189	0.148	0.199	-0.012	0.070	0.859
CCT*Age 16	0.201	0.143	0.153	-0.059	0.083	0.480
CCT*Age 17	0.122	0.151	0.416	-0.096	0.094	0.309
CCT*Age 18	0.057	0.163	0.728	-0.085	0.104	0.420
CCT*Age 19	0.046	0.167	0.782	0.024	0.118	0.841
CCT*Age 20+	0.040	0.177	0.820	0.039	0.125	0.753
CCT*Semi-rural	0.227	0.071	0.001	-0.085	0.049	0.085
CCT*Rural	0.346	0.095	0.000	-0.006	0.061	0.923
CCT*Highest grade	0.013	0.011	0.236	0.007	0.015	0.644
CCT*Asset index	0.021	0.012	0.067	-0.005	0.008	0.568
CTT*Never had sex	-0.056	0.063	0.372	-0.006	0.032	0.857
UCT*Age 14				-0.101	0.076	0.186
UCT*Age 15				-0.029	0.053	0.585
UCT*Age 16				-0.047	0.083	0.573
UCT*Age 17				-0.071	0.119	0.553
UCT*Age 18				-0.001	0.119	0.996
UCT*Age 19				-0.005	0.118	0.968
UCT*Age 20+				-0.011	0.133	0.936
UCT*Semi-rural				-0.057	0.049	0.253
UCT*Rural				0.788	0.057	.
UCT*Highest grade				-0.022	0.016	0.168
UCT*Asset index				0.018	0.008	0.017
UCT*Never had sex				-0.043	0.059	0.461
Sample Size	885			2,273		

Notes: Regressions are marginal effects probit models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. The outcome variable is an indicator for being successfully surveyed in Round 4. All interacted baseline covariates are demeaned prior to interaction so that the uninteracted treatment dummies can be interpreted as effects at the mean of all interacted covariates.

Table S3: Program impacts on educational qualifications and competencies (beneficiaries)

Panel A: Baseline Dropouts

	Educational Qualifications									Competencies (Standardized)					
	=1 if Passed Primary School (PSLC)			=1 if Passed Junior Secondary School (JCE)			=1 if Passed Secondary School (MSCE)			Fertilizer Application	Change Given	Sending a Text Message	Using a Calculator	Calculating Profits	Total Time Spent
	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	Two Years After Program					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
=1 if Conditional Schoolgirl	0.030	0.058**	0.081***	0.012	0.049**	0.034	0.004	0.003	0.016	-0.044	-0.014	0.101	0.065	0.094	-0.007
	(0.025)	(0.025)	(0.026)	(0.019)	(0.021)	(0.022)	(0.008)	(0.010)	(0.011)	(0.069)	(0.062)	(0.072)	(0.071)	(0.076)	(0.091)
Mean in Control Group	0.328	0.351	0.366	0.085	0.123	0.136	0.008	0.025	0.026	0.000	0.000	0.000	0.000	0.000	0.000
Sample Size	697	718	744	697	718	744	697	718	744	742	741	741	741	742	742
Panel B: Baseline Schoolgirls															
=1 if Conditional Schoolgirl	0.030	0.013	-0.014	-0.013	0.055*	0.033	-0.004*	0.005	0.006	0.015	0.048	0.077	0.060	-0.006	-0.113
	(0.039)	(0.024)	(0.019)	(0.022)	(0.028)	(0.028)	(0.002)	(0.011)	(0.021)	(0.071)	(0.071)	(0.070)	(0.054)	(0.076)	(0.085)
=1 if Unconditional Schoolgirl	0.046	0.030	0.017	0.002	0.016	0.010	-0.006*	-0.009	-0.065**	0.096	-0.017	0.161**	0.098	-0.045	-0.118
	(0.038)	(0.026)	(0.016)	(0.022)	(0.045)	(0.035)	(0.003)	(0.015)	(0.027)	(0.092)	(0.057)	(0.079)	(0.064)	(0.090)	(0.085)
p-value UCT vs. CCT	0.755	0.600	0.166	0.546	0.439	0.565	0.325	0.385	0.022	0.378	0.389	0.364	0.584	0.636	0.963
p-value Treatment	0.386	0.488	0.359	0.797	0.148	0.486	0.150	0.683	0.045	0.570	0.685	0.105	0.249	0.862	0.258
Mean in Control Group	0.496	0.776	0.879	0.144	0.337	0.537	0.004	0.054	0.170	0.000	0.000	0.000	0.000	0.000	0.000
Sample Size	1,967	2,019	2,047	1,967	2,019	2,047	1,967	2,019	2,047	2,048	2,046	2,047	2,047	2,048	2,048

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Competencies represent a set of skills that were anticipated to be sensitive to education and relevant for non-formal employment. The skills tested included reading and following instructions to apply fertilizer; making correct change during hypothetical market transactions; sending text messages and using the calculator on a mobile phone, and calculating profits under hypothetical business scenarios. All competency components are standardized to have a mean of zero and a standard deviation of one in the control group. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S4: Program impacts on sexual behavior (beneficiaries: **extensive** margin)

	=1 if Ever Had Sex			Number of Sexual Partners (lifetime)			=1 if Sexually Active During the Past 12 Months		
	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Baseline Dropouts									
=1 if Conditional Schoolgirl	-0.036*	-0.034	-0.004	0.004	-0.118	-0.023	-0.123***	-0.094**	-0.046
	(0.020)	(0.021)	(0.010)	(0.153)	(0.153)	(0.095)	(0.035)	(0.037)	(0.028)
Mean in Control Group	0.814	0.918	0.976	1.395	1.734	2.063	0.503	0.674	0.830
Sample Size	698	718	744	698	718	744	697	718	744
Panel B: Baseline Schoolgirls									
=1 if Conditional Schoolgirl	-0.009	-0.003	0.005	-0.023	0.005	0.005	-0.009	0.001	-0.030
	(0.017)	(0.024)	(0.035)	(0.040)	(0.048)	(0.061)	(0.023)	(0.029)	(0.035)
=1 if Unconditional Schoolgirl	-0.022	0.003	0.041	-0.044	-0.007	0.108	-0.021	-0.036	0.037
	(0.021)	(0.030)	(0.036)	(0.049)	(0.036)	(0.066)	(0.030)	(0.032)	(0.044)
p-value UCT vs. CCT	0.581	0.864	0.414	0.699	0.815	0.142	0.728	0.327	0.177
p-value Treatment	0.551	0.984	0.519	0.627	0.969	0.218	0.768	0.514	0.395
Mean in Control Group	0.303	0.455	0.701	0.335	0.559	1.045	0.175	0.308	0.563
Sample Size	1,965	2,016	2,048	1,964	2,016	2,047	1,965	2,015	2,048

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. We correct for inconsistencies in 'ever had sex' across rounds. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S5: Program impacts on sexual behavior (beneficiaries: **intensive** margin)

	Age at First Sex			=1 if Older Partner			=1 if Use a Condom	
	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	End of Program	Two Years After Program
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Baseline Dropouts								
=1 if Conditional Schoolgirl	-0.064	-0.061	0.110	0.018	-0.005	0.015	0.046	0.030
	(0.137)	(0.144)	(0.133)	(0.054)	(0.045)	(0.037)	(0.037)	(0.030)
Mean in Control Group	16.250	16.578	16.790	0.230	0.300	0.309	0.159	0.156
Sample Size	525	625	723	303	427	578	446	600
Panel B: Baseline Schoolgirls								
=1 if Conditional Schoolgirl	0.220	0.136	0.147	-0.074	-0.006	-0.041	-0.006	0.015
	(0.146)	(0.130)	(0.146)	(0.050)	(0.044)	(0.038)	(0.055)	(0.041)
=1 if Unconditional Schoolgirl	-0.152	-0.039	-0.207	0.022	-0.081	0.018	0.102	0.057
	(0.179)	(0.189)	(0.127)	(0.103)	(0.057)	(0.049)	(0.086)	(0.048)
p-value UCT vs. CCT	0.064	0.404	0.052	0.351	0.258	0.248	0.268	0.482
p-value Treatment	0.143	0.536	0.128	0.291	0.367	0.422	0.483	0.479
Mean in Control Group	15.731	16.393	17.199	0.193	0.274	0.304	0.247	0.268
Sample Size	522	893	1,494	376	661	1,162	672	1,183

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). We correct for inconsistencies in 'ever had sex' across rounds. 'Age at First Sex' is defined for those that had ever had sex. 'Older Partner' is defined as having a partner who is 5 years older or more in the past 12 months. 'Condom Use' is defined as using a condom at last sex with most recent sexual partner. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S6: Program impacts on health and nutrition (beneficiaries)

	=1 if Suffers from Psychological Distress			Number of Times Respondent Ate Protein Rich Foods During the Past 7 Days (out of 21)		
	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Baseline Dropouts						
=1 if Conditional Schoolgirl	-0.002	0.010	0.038	0.326	0.224	0.228
	(0.039)	(0.036)	(0.042)	(0.202)	(0.192)	(0.181)
Mean in Control Group	0.463	0.314	0.424	3.678	3.989	3.741
Sample Size	698	715	743	698	718	744
Panel B: Baseline Schoolgirls						
=1 if Conditional Schoolgirl	-0.068**	-0.037	-0.030	0.385**	0.596***	0.072
	(0.032)	(0.047)	(0.032)	(0.195)	(0.174)	(0.141)
=1 if Unconditional Schoolgirl	-0.139***	-0.026	-0.002	0.445**	0.338**	-0.043
	(0.035)	(0.054)	(0.046)	(0.199)	(0.153)	(0.240)
p-value UCT vs. CCT	0.068	0.860	0.552	0.814	0.215	0.672
p-value Treatment	0.000	0.677	0.627	0.023	0.001	0.858
Mean in Control Group	0.372	0.313	0.369	3.967	4.052	4.134
Sample Size	1,963	2,013	2,045	1,967	2,018	2,047

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Psychological distress is equal to one if the summed GHQ- 12 score is equal to three or higher, and is zero otherwise. Protein rich foods are defined as those containing animal proteins, i.e. meat, fish, and eggs. The number of days each item was consumed over the past week are summed to create the outcome variable. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S7: Program impacts on labor market outcomes and consumption (beneficiaries: **secondary** outcomes)

	Effective Daily Wage (Past 7 Days) 2012 USD	Labor Income (Past 5 Seasons) 2012 USD	=1 if Any Wage Work in Past 3 Months	Real Total Household Monthly Consumption (USD)		
	Two Years After Program			During Program	End of Program	Two Years After Program
Panel A: Baseline Dropouts						
=1 if Conditional Schoolgirl	-0.228	4.129	-0.020	-0.257	-1.941*	0.535
	(0.148)	(8.620)	(0.037)	(1.029)	(1.113)	(1.130)
Mean in Control Group	0.753	52.840	0.366	17.502	20.860	17.977
Sample Size	263	744	744	712	719	737
Panel B: Baseline Schoolgirls						
=1 if Conditional Schoolgirl	0.121	7.476	-0.010	3.192**	3.223**	2.804*
	(0.424)	(7.466)	(0.030)	(1.261)	(1.364)	(1.432)
=1 if Unconditional Schoolgirl	-0.549*	10.688	0.001	-0.586	-0.880	-0.817
	(0.285)	(12.721)	(0.055)	(1.441)	(1.524)	(1.876)
p-value UCT vs. CCT	0.278	0.829	0.838	0.032	0.034	0.127
p-value Treatment	0.121	0.420	0.939	0.030	0.044	0.137
Mean in Control Group	0.902	33.302	0.250	18.638	23.342	20.774
Sample Size	465	2,049	2,049	2,006	2,021	2,040

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Effective daily wage is calculated using earnings and activities in the past seven days. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S8: Program impacts on empowerment (beneficiaries: **secondary** outcomes)

Panel A: Baseline Dropouts	Empowerment			
	Index of Self-Efficacy (Standardized)	Index of Preferences for Child Education (Standardized)	Index of Social Participation (Standardized)	Aspirations
	(1)	(2)	(3)	(4)
=1 if Conditional Schoolgirl	-0.041	-0.020	-0.052	-0.221
	(0.076)	(0.079)	(0.068)	(0.225)
Mean in Control Group	0.000	0.000	0.000	3.267
Sample Size	744	744	744	744
Panel B: Baseline Schoolgirls				
=1 if Conditional Schoolgirl	0.059	-0.004	-0.026	0.235
	(0.079)	(0.076)	(0.068)	(0.228)
=1 if Unconditional Schoolgirl	-0.149	-0.106	-0.095	0.004
	(0.100)	(0.087)	(0.069)	(0.207)
p-value UCT vs. CCT	0.061	0.343	0.424	0.379
p-value Treatment	0.170	0.477	0.393	0.566
Mean in Control Group	0.000	0.000	0.000	3.352
Sample Size	2,049	2,049	2,049	2,049

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Detail on the construction of the indices can be found at https://drive.google.com/file/d/1hvI79ltywocFr-pafqz8_Dtg2ZXNhcHd/view. Aspirations asks the respondent where she sees herself on a 10-step ladder comparing today to five years from now, where zero represents the worst possible life she could have and 10 represents the best possible life she could have. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S9: Program impacts on key outcome using full Round 2 and Round 3 data (*baseline dropouts*)

Panel A: Round 2 Impacts	Highest Grade Attended		=1 if Ever Married		=1 if Ever Pregnant		Number of Live Births		=1 if HIV Positive	
	Round 2 Sample	Round 4 Sample	Round 2 Sample	Round 4 Sample	Round 2 Sample	Round 4 Sample	Round 2 Sample	Round 4 Sample	Round 2 Sample	Round 4 Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(8)	(9)	(10)	(11)
=1 if Conditional Schoolgirl	0.546***	0.579***	-0.102***	-0.140***	-0.053*	-0.057*	-0.003	-0.005	0.032	0.022
	(0.068)	(0.073)	(0.026)	(0.029)	(0.029)	(0.030)	(0.033)	(0.033)	(0.027)	(0.024)
Mean in Control Group	6.462	6.345	0.278	0.291	0.607	0.610	0.521	0.520	0.082	0.060
Sample Size	800	697	801	698	801	698	801	698	417	373
Panel A: Round 3 Impacts										
	Round 3 Sample	Round 4 Sample	Round 3 Sample	Round 4 Sample	Round 3 Sample	Round 4 Sample	Round 3 Sample	Round 4 Sample	Round 3 Sample	Round 4 Sample
=1 if Conditional Schoolgirl	0.522***	0.558***	-0.139***	-0.157***	-0.079***	-0.081***	-0.097**	-0.095**	0.026	0.020
	(0.099)	(0.102)	(0.036)	(0.037)	(0.026)	(0.027)	(0.042)	(0.044)	(0.025)	(0.023)
Mean in Control Group	7.078	6.967	0.571	0.575	0.783	0.784	0.817	0.819	0.105	0.094
Sample Size	801	718	801	718	801	718	801	718	763	694

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. We correct for inconsistencies in 'ever married' and 'ever pregnant' across rounds. All regressions are weighted to make them representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S10: Program impacts on key outcome using full Round 2 and Round 3 data (*baseline schoolgirls*)

	Highest Grade Attended		=1 if Ever Married		=1 if Ever Pregnant		Number of Live Births		=1 if HIV Positive	
Panel A: Round 2 Impacts										
	Round 2 Sample	Round 4 Sample	Round 2 Sample	Round 4 Sample	Round 2 Sample	Round 4 Sample	Round 2 Sample	Round 4 Sample	Round 2 Sample	Round 4 Sample
	(1)	(2)	(3)	(4)	(5)	(6)	(8)	(9)	(10)	(11)
=1 if Conditional Schoolgirl	0.081	0.078	0.006	0.000	0.014	0.008	0.019	0.023*	-0.020**	-0.020**
	(0.089)	(0.090)	(0.012)	(0.012)	(0.014)	(0.015)	(0.014)	(0.014)	(0.009)	(0.009)
=1 if Unconditional Schoolgirl	0.134	0.122	-0.030**	-0.033***	-0.012	-0.013	0.013	0.013	-0.018	-0.015
	(0.107)	(0.109)	(0.012)	(0.012)	(0.017)	(0.017)	(0.017)	(0.017)	(0.012)	(0.012)
p-value UCT vs. CCT	0.644	0.708	0.014	0.026	0.194	0.314	0.773	0.641	(0.818)	0.616
p-value Treatment	0.407	0.469	0.022	0.023	0.404	0.600	0.302	0.209	(0.080)	0.112
Mean in Control Group	8.597	8.590	0.047	0.047	0.092	0.092	0.056	0.055	0.030	0.026
Sample Size	2,146	1,965	2149	1,967	2,148	1,966	2,148	1,966	1,287	1,192
Panel A: Round 3 Impacts										
	Round 3 Sample	Round 4 Sample	Round 3 Sample	Round 4 Sample	Round 3 Sample	Round 4 Sample	Round 3 Sample	Round 4 Sample	Round 3 Sample	Round 4 Sample
=1 if Conditional Schoolgirl	0.102	0.126*	-0.008	-0.010	0.035	0.027	0.006	0.003	-0.004	-0.003
	(0.072)	(0.069)	(0.023)	(0.024)	(0.027)	(0.027)	(0.021)	(0.022)	(0.011)	(0.011)
=1 if Unconditional Schoolgirl	0.099	0.103	-0.075***	-0.083***	-0.059**	-0.063**	-0.050*	-0.055*	-0.020*	-0.019*
	(0.117)	(0.121)	(0.023)	(0.024)	(0.027)	(0.028)	(0.027)	(0.030)	(0.012)	(0.012)
p-value UCT vs. CCT	0.975	0.854	0.024	0.018	0.006	0.009	(0.069)	0.075	0.243	0.237
p-value Treatment	0.323	0.174	0.005	0.004	0.019	0.025	(0.148)	0.151	0.247	0.249
Mean in Control Group	9.699	9.677	0.180	0.180	0.248	0.247	0.201	0.199	0.037	0.035
Sample Size	2,176	2,019	2,175	2,018	2,176	2,019	2,176	2,019	2,136	2,002

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. We correct for inconsistencies in 'ever married' and 'ever pregnant' across rounds. All regressions are weighted to make them representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S11: Attrition Analysis for Primary Round 4 Outcomes (Baseline Dropouts)

	Dependent Variable: =1 if Conditional Schoolgirl							
	Lower Bounds:			Unadjusted	IPW	Upper Bounds		
Outcomes	(-) .25 SD	(-) .1 SD	Lee			Lee	(+) .1 SD	(+) .25 SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Highest Grade Completed	0.420***	0.561***	0.616***	0.621***	0.606***	0.632***	0.750***	0.891***
	(0.116)	(0.114)	(0.125)	(0.125)	(0.128)	(0.125)	(0.113)	(0.114)
Competencies Score (Standardized)	-0.025	0.024	0.061	0.064	0.065	0.089	0.090*	0.139***
	(0.048)	(0.047)	(0.057)	(0.057)	(0.057)	(0.056)	(0.047)	(0.047)
=1 if Ever Married	-0.138***	-0.119***	-0.106***	-0.107***	-0.108***	-0.112***	-0.093***	-0.073***
	(0.027)	(0.027)	(0.032)	(0.032)	(0.033)	(0.032)	(0.027)	(0.027)
=1 if Ever Pregnant	-0.059***	-0.045***	-0.040*	-0.040*	-0.041**	-0.044**	-0.026	-0.012
	(0.018)	(0.017)	(0.021)	(0.021)	(0.021)	(0.021)	(0.017)	(0.017)
Number of Live Births	-0.199***	-0.162***	-0.127**	-0.147***	-0.145***	-0.153***	-0.113**	-0.076*
	(0.044)	(0.044)	(0.052)	(0.054)	(0.055)	(0.053)	(0.044)	(0.045)
= if HIV Positive	-0.021	-0.002	0.019	0.012	0.007	0.011	0.023	0.041*
	(0.024)	(0.024)	(0.026)	(0.026)	(0.026)	(0.026)	(0.024)	(0.024)
=1 if Anemic	-0.002	0.024	0.045	0.039	0.041	0.038	0.059**	0.086***
	(0.029)	(0.028)	(0.035)	(0.035)	(0.035)	(0.035)	(0.028)	(0.028)
Opportunity Cost of Time	-0.146**	-0.084	-0.038	-0.037	-0.041	0.065	-0.002	0.060
	(0.066)	(0.065)	(0.079)	(0.079)	(0.079)	(0.052)	(0.064)	(0.064)
Typical Daily Wage in Last Three Months	-0.216***	-0.173***	-0.141**	-0.140**	-0.132*	-0.074	-0.115**	-0.071
	(0.055)	(0.054)	(0.068)	(0.068)	(0.069)	(0.055)	(0.054)	(0.055)
Proportion of Hours Spent in Self-Employment or Paid Work in Past	-0.020***	-0.015**	-0.011	-0.011	-0.012	-0.007	-0.008	-0.003
	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.008)	(0.007)	(0.007)
Super Index of Overall Empowerment (Standardized)	-0.175***	-0.128**	-0.097	-0.083	-0.073	-0.065	-0.065	-0.017
	(0.063)	(0.062)	(0.074)	(0.074)	(0.075)	(0.072)	(0.061)	(0.061)
Change in Subjective Wellbeing from Five Years Ago to Today	-0.319	-0.157	-0.091	-0.032	-0.050	0.048	0.059	0.221
	(0.200)	(0.198)	(0.232)	(0.232)	(0.235)	(0.226)	(0.196)	(0.196)

Notes: Regressions examine extensive margin outcomes using OLS models with robust standard errors clustered at the EA level. Each coefficient in this table comes from a different regression of the outcome on a dummy for treatment. Column (4) re-presents the main results using sampling weights. Column (5) weights by attrition propensity weights, using the standard battery of covariates and their interactions with the treatment to predict attrition. Columns (3) and (6) provide the lower and upper bounds for the treatment effects using Lee Bounds. Columns (1), (2), (7), and (8) use survey weights and impute missing outcomes by taking the treatment-arm specific mean and adding (subtracting) an amount equal to .1 or .25 of the standard deviation of the outcome in that arm. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, an indicator for never had sex, and whether the respondent participated in the pilot phase of the development of the testing instruments. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S12: Attrition Analysis for Primary Round 4 Outcomes (Baseline Schoolgirls: CCT vs. Control/UCT vs. Control)

	Dependent Variable: =1 if Conditional Schoolgirl								Dependent Variable: =1 if Unconditional Schoolgirl							
	Lower Bounds:			Un-adjusted	IPW	Upper Bounds			Lower Bounds:			Un-adjusted	IPW	Upper Bounds		
Outcomes	(-) .25 SD	(-) .1 SD	Lee			Lee	(+) .1 SD	(+) .25 SD	(-) .25 SD	(-) .1 SD	Lee			Lee	(+) .1 SD	(+) .25 SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Highest Grade Completed	0.049	0.103	0.041	0.120	0.122	0.208***	0.174**	0.228***	0.055	0.106	0.052	0.095	0.095	0.210*	0.175	0.226**
	(0.075)	(0.075)	(0.084)	(0.080)	(0.080)	(0.075)	(0.075)	(0.076)	(0.112)	(0.112)	(0.132)	(0.129)	(0.129)	(0.123)	(0.113)	(0.115)
Competencies Score (Standardized)	0.031	0.060	-0.045	0.065	0.066	0.129**	0.099*	0.129**	0.077	0.106*	0.028	0.098	0.101	0.176***	0.144**	0.173***
	(0.056)	(0.056)	(0.050)	(0.058)	(0.058)	(0.060)	(0.056)	(0.056)	(0.061)	(0.060)	(0.067)	(0.067)	(0.068)	(0.063)	(0.059)	(0.058)
=1 if Ever Married	-0.068**	-0.054**	-0.065**	-0.035	-0.035	-0.018	-0.035	-0.020	-0.047	-0.033	-0.049	-0.010	-0.010	0.010	-0.014	-0.000
	(0.027)	(0.027)	(0.027)	(0.027)	(0.028)	(0.027)	(0.027)	(0.027)	(0.044)	(0.044)	(0.045)	(0.046)	(0.048)	(0.047)	(0.043)	(0.043)
=1 if Ever Pregnant	-0.056*	-0.041	-0.046	-0.024	-0.025	-0.001	-0.022	-0.007	-0.036	-0.021	-0.034	-0.001	0.000	0.027	-0.002	0.013
	(0.034)	(0.033)	(0.035)	(0.034)	(0.034)	(0.033)	(0.033)	(0.033)	(0.040)	(0.040)	(0.042)	(0.042)	(0.043)	(0.039)	(0.040)	(0.040)
Number of Live Births	-0.020	-0.001	-0.027	0.020	0.020	0.044	0.025	0.044	-0.067	-0.049	-0.094*	-0.024	-0.022	0.002	-0.024	-0.006
	(0.036)	(0.036)	(0.035)	(0.036)	(0.037)	(0.036)	(0.036)	(0.036)	(0.045)	(0.044)	(0.049)	(0.046)	(0.048)	(0.046)	(0.043)	(0.043)
= if HIV Positive	-0.013	-0.005	-0.057***	-0.001	-0.002	0.002	0.005	0.013	-0.022	-0.012	-0.055***	-0.002	-0.002	0.001	0.000	0.010
	(0.018)	(0.018)	(0.008)	(0.019)	(0.020)	(0.020)	(0.018)	(0.018)	(0.021)	(0.021)	(0.008)	(0.023)	(0.023)	(0.025)	(0.021)	(0.021)
=1 if Anemic	-0.020	-0.005	-0.031	0.012	0.011	0.029	0.016	0.031	-0.093***	-0.076***	-0.105***	-0.065*	-0.067**	-0.054	-0.053*	-0.036
	(0.027)	(0.027)	(0.033)	(0.031)	(0.031)	(0.031)	(0.027)	(0.028)	(0.029)	(0.029)	(0.031)	(0.033)	(0.033)	(0.035)	(0.029)	(0.029)
Opportunity Cost of Time	-0.166*	-0.105	-0.343***	-0.051	-0.049	-0.015	-0.022	0.039	-0.208***	-0.152**	-0.320***	-0.115	-0.112	-0.083	-0.078	-0.022
	(0.087)	(0.086)	(0.060)	(0.101)	(0.100)	(0.102)	(0.086)	(0.087)	(0.068)	(0.067)	(0.062)	(0.074)	(0.071)	(0.079)	(0.065)	(0.064)
Typical Daily Wage in Last Three	-0.048	-0.026	-0.131***	-0.011	-0.009	-0.003	0.005	0.027	-0.008	0.016	-0.177***	0.036	0.036	0.052	0.048	0.072
Months	(0.056)	(0.055)	(0.035)	(0.058)	(0.058)	(0.061)	(0.054)	(0.053)	(0.097)	(0.097)	(0.038)	(0.104)	(0.105)	(0.109)	(0.097)	(0.097)
Proportion of Hours Spent in Self-	-0.001	0.002	-0.010***	0.003	0.004	0.005	0.005	0.007	-0.003	-0.000	-0.021***	0.002	0.001	0.004	0.003	0.005
Employment or Paid Work in Past	(0.005)	(0.005)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.003)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
Super Index of Overall	-0.007	0.023	-0.079	0.049	0.046	0.121	0.063	0.093	-0.199***	-0.168**	-0.260***	-0.159*	-0.156*	-0.079	-0.127*	-0.097
Empowerment (Standardized)	(0.076)	(0.076)	(0.072)	(0.082)	(0.082)	(0.085)	(0.075)	(0.075)	(0.075)	(0.075)	(0.071)	(0.081)	(0.082)	(0.089)	(0.075)	(0.075)
Change in Subjective Wellbeing	0.101	0.189	-0.142	0.276	0.275	0.706***	0.306*	0.394**	0.019	0.104	-0.108	0.176	0.174	0.515***	0.217	0.301*
from Five Years Ago to Today	(0.168)	(0.169)	(0.157)	(0.187)	(0.189)	(0.173)	(0.172)	(0.174)	(0.175)	(0.174)	(0.190)	(0.190)	(0.190)	(0.179)	(0.172)	(0.172)

Notes: Regressions examine extensive margin outcomes using OLS with robust standard errors clustered at the EA level. Each coefficient in this table comes from a different regression of the outcome on a dummy for treatment. Column (4) re-presents the main results using sampling weights. Column (5) weights by attrition propensity weights, using the standard battery of covariates and their interactions with the treatment to predict attrition. Columns (3) and (6) provide the lower and upper bounds for the treatment effects using Lee Bounds. Columns (1), (2), (7), and (8) use survey weights and impute missing outcomes by taking the treatment-arm specific mean and adding (subtracting) an amount equal to .1 or .25 of the standard deviation of the outcome in that arm. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, an indicator for never had sex, and whether the respondent participated in the pilot phase of the development of the testing instruments. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S13: Attrition Analysis for Primary Round 4 Outcomes (Baseline Schoolgirls: CCT vs. UCT)

	Dependent Variable: =1 if Unconditional Schoolgirl							
	Lower Bounds:			Un-adjusted	IPW	Upper Bounds		
Outcomes	(-) .25 SD	(-) .1 SD	Lee			Lee	(+) .1 SD	(+) .25 SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Highest Grade Completed	-0.058	-0.022	-0.027	-0.021	-0.020	-0.005	0.026	0.062
	(0.116)	(0.117)	(0.129)	(0.129)	(0.129)	(0.127)	(0.119)	(0.120)
Competencies Score (Standardized)	0.009	0.029	0.015	0.029	0.032	0.039	0.057	0.078
	(0.069)	(0.068)	(0.072)	(0.072)	(0.073)	(0.070)	(0.067)	(0.067)
=1 if Ever Married	-0.013	-0.003	0.013	0.011	0.009	0.006	0.010	0.021
	(0.048)	(0.047)	(0.049)	(0.049)	(0.050)	(0.049)	(0.046)	(0.045)
=1 if Ever Pregnant	-0.013	-0.003	0.013	0.010	0.010	0.006	0.011	0.022
	(0.045)	(0.044)	(0.046)	(0.046)	(0.047)	(0.047)	(0.044)	(0.044)
Number of Live Births	-0.083	-0.069	-0.051	-0.053	-0.052	-0.061	-0.051	-0.037
	(0.051)	(0.050)	(0.054)	(0.053)	(0.054)	(0.053)	(0.050)	(0.049)
= if HIV Positive	-0.015	-0.007	0.004	0.003	0.003	-0.006	0.004	0.011
	(0.027)	(0.027)	(0.029)	(0.029)	(0.029)	(0.028)	(0.027)	(0.027)
=1 if Anemic	-0.093**	-0.078**	-0.072*	-0.074*	-0.075*	-0.086**	-0.058	-0.043
	(0.037)	(0.037)	(0.041)	(0.042)	(0.041)	(0.042)	(0.037)	(0.037)
Opportunity Cost of Time	-0.084	-0.047	-0.119	-0.040	-0.035	-0.037	0.002	0.040
	(0.093)	(0.091)	(0.095)	(0.106)	(0.103)	(0.107)	(0.089)	(0.088)
Typical Daily Wage in Last Three	0.050	0.063	-0.058	0.079	0.077	0.080	0.082	0.095
Months	(0.102)	(0.102)	(0.046)	(0.108)	(0.108)	(0.109)	(0.101)	(0.101)
Proportion of Hours Spent in Self-	-0.006	-0.004	-0.013*	-0.003	-0.003	-0.003	-0.002	0.000
Employment or Paid Work in Past	(0.009)	(0.009)	(0.007)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)
Super Index of Overall	-0.241**	-0.219**	-0.243**	-0.221**	-0.214**	-0.206*	-0.189*	-0.167*
Empowerment (Standardized)	(0.097)	(0.098)	(0.104)	(0.106)	(0.107)	(0.108)	(0.099)	(0.101)
Change in Subjective Wellbeing	-0.221	-0.160	-0.203	-0.134	-0.134	-0.073	-0.079	-0.017
from Five Years Ago to Today	(0.207)	(0.206)	(0.219)	(0.221)	(0.221)	(0.202)	(0.205)	(0.205)

Notes: Regressions examine extensive margin outcomes using OLS models with robust standard errors clustered at the EA level. Each coefficient in this table comes from a different regression of the outcome on a dummy for treatment. Column (4) re-presents the main results using sampling weights. Column (5) weights by attrition propensity weights, using the standard battery of covariates and their interactions with the treatment to predict attrition. Columns (3) and (6) provide the lower and upper bounds for the treatment effects using Lee Bounds. Columns (1), (2), (7), and (8) use survey weights and impute missing outcomes by taking the treatment-arm specific mean and adding (subtracting) an amount equal to .1 or .25 of the standard deviation of the outcome in that arm. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, an indicator for never had sex, and whether the respondent participated in the pilot phase of the development of the testing instruments. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S14: Heterogeneity of Child Height Mediators by Transfer Amount

	Core Respondent Outcomes					
	Mentally Ill			**Meals Eaten**		
	R2	R3	R4	R2	R3	R4
	(1)	(2)	(3)	(4)	(5)	(6)
=1 if Conditional Schoolgirl	-0.132**	0.009	-0.073	0.462	0.975***	0.527*
(minimum transfer amount)	(0.054)	(0.085)	(0.069)	(0.412)	(0.264)	(0.294)
Household transfer amount, CCT	0.031***	0.002	-0.001	-0.142*	-0.056	-0.060
	(0.009)	(0.013)	(0.011)	(0.084)	(0.080)	(0.051)
Individual transfer amount, CCT	-0.009	-0.027	0.024	0.212*	-0.117	-0.134
	(0.018)	(0.020)	(0.027)	(0.114)	(0.081)	(0.121)
=1 if Unconditional Schoolgirl	-0.118**	-0.010	-0.007	0.392	-0.144	0.161
(minimum transfer amount)	(0.060)	(0.076)	(0.069)	(0.371)	(0.357)	(0.387)
Household transfer amount, UCT	0.009	0.018	0.029	-0.001	0.030	-0.120
	(0.015)	(0.019)	(0.020)	(0.065)	(0.059)	(0.083)
Individual transfer amount, UCT	-0.023	-0.041**	-0.039***	0.013	0.175	0.071
	(0.020)	(0.019)	(0.015)	(0.106)	(0.112)	(0.126)
Number of observations	2,145	2,170	2,045	2,149	2,175	2,047

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Psychological distress is equal to one if the summed GHQ- 12 score is equal to three or higher, and is zero otherwise. Protein rich foods are defined as those containing animal proteins, i.e. meat, fish, and eggs. The number of days each item was consumed over the past week are summed to create the outcome variable. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.